

OM of: US-09-303-518d-465 to: PIR_71.* out_format : pfs

Date: Jun 30, 2002 7:47 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame-n2p.model -DEV=x1h
-Q/c9n2.1/USPTO_spool/US09303518/runat_28062002.142713.4317/app_query.fasta.1.23501
-DB=PIR_71 -QFMT=fastan -SUFFIX=ipr -GAPOP=12.000 -GAPEXT=4.500
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.500 -GAPOP=6.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DEL0P=6.000
-RAXPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DEL0P=6.000
-DELETE=7.000 -START=1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=100 -DOCALLIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0
-ALIGN=45 -MODE=LOCAL -OUTFMT=pls -NORM=ext -HEA51E=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09303518_c9n1.1.1092
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-303-518d-465

Query length: 1671

Database: PIR_71.*

Database sequences: 283138

Search length: 96089334

Search time (sec): 504.980000

score list:

Sequence	Strd Orig	zScore	Escore	len	Documentation
PIR2:H81782	730.50	895.53	1.7e-42	498	adhesin MafB2 NMA2113 [imported]
PIR2:B82028	419.50	510.65	4.8e-21	482	adhesin NMA0324 [imported] - Ne
PIR2:T48814	170.00	188.75	0.0010	1952	hypothetical protein 1566.220
PIR2:T45025	159.50	170.04	0.0062	3570	mucin MUC5B, tracheobronchial
PIR2:T30826	156.50	170.96	0.0090	2187	nascent polypeptide-associated
PIR2:A4932	150.50	164.20	0.0155	3020	mucin 2 precursor, intestinal
PIR2:T47141	150.50	176.99	0.0172	528	gastric mucin (clone PGW-2)
PIR2:T46134	150.50	176.97	0.0172	528	hypothetical protein [imported]
PIR2:S53047	149.50	175.49	0.0203	543	mucin J0U7 - human
PIR2:T22808	149.50	171.56	0.0237	770	hypothetical protein F56H9.1 -
PIR2:T15316	148.50	166.56	0.0283	1223	hypothetical protein B0302.1 -
PIR2:S48478	147.50	165.65	0.0368	1367	glucan 1,4-alpha-glucosidase 0
PIR2:T29634	146.50	167.81	0.0358	825	hypothetical protein C12D12.1 -
PIR2:S57180	141.00	157.77	0.0922	1161	probable membrane protein YMR
PIR2:T43263	140.50	159.35	0.0950	920	cell division protein dmfl/mid
PIR2:T30345	140.00	154.93	0.1121	1375	dextranucrase (EC 2.4.1.5) pr
PIR2:J00405	139.00	156.37	0.1158	1106	hypothetical protein 119.5k protein (u
PIR2:A48018	139.00	165.94	0.0996	377	mucin 7 precursor, salivary - h
PIR2:T00345	139.00	151.19	0.1390	1791	hypothetical protein KIAA0324
PIR2:T39758	137.50	159.45	0.1403	615	proline-serine rich protein - h
PIR2:T31422	137.00	150.57	0.1831	1473	C-terminal domain-binding prot
PIR2:T34230	136.50	155.53	0.1747	817	verprolin - yeast (Saccharomyce
PIR2:S51342	135.50	153.39	0.2868	3942	Baasoon protein - mouse
PIR2:T42730	135.00	159.59	0.1939	437	hypothetical protein 3 (gyrB re
PIR2:C39135	134.00	148.92	0.2813	1184	atrophin-1 - human
PIR2:G01763	134.00	148.22	0.2858	1275	hypothetical protein H02F09.3
PIR2:T33369	133.50	144.32	0.3186	1460	immediate-early protein It180
PIR2:EDBE1F	133.00	144.66	0.3511	1630	ascites statoglycoprotein 16180
PIR2:A53577	132.50	146.75	0.2620	1299	mucin 5AC (clone JER58) - human
PIR2:S53363	132.50	146.19	0.3642	1299	hypothetical protein DKFZ4434V
PIR2:T47182	132.00	148.57	0.4116	1589	hypothetical protein B7F21.40
PIR2:T13606	131.50	146.66	0.3932	886	glycoprotein 350/220 - human he
PIR2:S29605	131.50	145.83	0.4184	1184	atrophin-1 - human
PIR2:S50832	130.50	140.65	0.5360	1794	hypothetical divergent repeat
PIR2:T384593	130.00	144.66	0.4996	891	probable proline-rich protein
PIR2:G84953	130.00	144.48	0.5249	1122	exit protein - Mycobacterium s
PIR2:T14180	130.00	136.35	0.6308	2649	hypothetical protein B7F21.40
PIR2:T51023	129.50	144.01	0.4800	5104	hypothetical protein T620.190
PIR2:T12970	129.50	144.01	0.5663	1104	probable membrane protein YMR3
PIR2:S59310	129.00	152.96	0.4938	402	hypothetical protein [imported]

PIR2:T49415	129.00	148.64	0.5444	634	hypothetical protein BID4.30
PIR2:S24169	128.50	155.36	0.4993	792	mucin - rat
PIR1:VGBEX1	128.50	145.86	0.6190	797	glycoprotein X precursor - e
PIR2:J0W067	128.00	147.02	0.6436	660	chitinase (EC 3.2.1.14) A -
PIR2:S25345	128.00	138.59	0.7789	1609	probable membrane protein Y
PIR2:T21700	127.50	144.06	0.7348	846	hypothetical protein F35E2.6
PIR2:T49988	127.00	147.24	0.7299	566	ovule development protein-11
PIR2:T37455	127.00	140.66	0.8483	1142	enamelin precursor - pig
PIR2:T31421	127.00	140.34	0.8532	1173	C-terminal domain-binding P
PIR2:S63378	126.50	142.31	0.8733	893	hypothetical protein YNC047W
PIR2:T45462	126.00	141.98	0.9371	866	membrane glycoprotein [import
PIR2:T18300	126.00	140.15	0.9765	1050	serine/threonine-specific P
PIR2:T17415	126.00	138.87	1.01	1203	myeloid surface antigen CS
PIR2:S10099	125.50	145.12	0.9318	582	transcription factor Irf-1 -
PIR2:T26517	125.00	134.73	1.26	1634	hypothetical protein Y18D10
PIR2:T33816	124.50	143.01	1.11	658	transcription regulator Pan-
PIR2:T14181	124.50	134.08	1.36	1639	peptide synthetase homolog
PIR2:T54463	124.00	139.49	1.29	867	membrane glycoprotein [import
PIR2:T44825	124.00	139.13	1.30	901	phosphoprotein, synapse-spec
PIR2:T18535	124.00	136.81	1.37	1151	high molecular mass nuclear
PIR2:T19964	124.00	132.31	1.44	1023	hypothetical protein C46C2.
PIR2:T48997	123.50	137.31	1.58	15281	epsin-like protein - Arabid
PIR2:S41309	123.00	138.31	1.51	862	cyclosporin synthetase - c
PIR2:T46289	123.00	137.83	1.52	907	membrane antigen gp350 - hum
PIR1:Q0BE21	123.00	132.93	2.09	1611	hypothetical protein SPAC23
PIR2:T38236	123.00	123.93	2.12	3338	Baasoon protein - rat
PIR2:T42761	122.50	126.17	2.12	2910	otogelin - mouse
PIR2:T42214	122.00	142.13	1.58	505	mucin 6, gastric (3'-repeat c
PIR2:S46629	122.00	140.82	1.62	580	probable mucin DKFZp34C0196.
PIR2:T43481	122.00	138.41	1.93	1296	collar protein isoform C.
PIR2:T13936	122.00	128.40	2.15	2154	hypothetical protein PA625
PIR2:F83068	121.50	136.08	1.93	896	clathrin assembly protein AP
PIR2:S34326	121.50	134.75	1.99	1032	hypothetical protein K06A9.
PIR2:T50144	121.50	127.44	2.34	2232	hypothetical protein K06A9.
PIR2:S55049	121.00	139.10	1.82	610	early growth response protei
PIR2:T55049	121.00	138.91	1.92	623	osteoblast specific factor 2
PIR2:T55049	121.00	137.68	1.93	709	Iga-specific metalloendopept
PIR2:F75523	121.00	137.68	1.99	2468	hypothetical protein PA1874
PIR2:G64057	121.00	125.87	2.59	332	probable rhoA protein - Myco
PIR2:A83412	120.50	144.24	1.83	477	cell surface glycoprotein - hu
PIR2:D06065	120.50	140.81	1.98	674	mucin 5AC (clone JER47) - hu
PIR2:S53362	120.50	137.54	2.13	851	hypothetical protein F5511.
PIR2:A69111	120.50	135.33	2.24	1224	microtubule-associated prot
PIR2:T22656	120.50	131.89	2.62	1777	hypothetical protein T19D12
PIR2:T14309	120.00	128.36	2.42	632	5'-nucleotidase [imported] -
PIR2:T40353	120.00	137.53	2.27	636	probable envelope protein -
PIR2:C98264	120.00	137.47	2.27	750	probable envelope protein -
PIR2:AR3020	120.00	135.91	2.36	968	hypothetical protein KIAA069
PIR2:T42614	120.00	121.31	3.28	3507	hypothetical protein ZK783.
PIR2:T00353	120.00	135.50	2.49	753	OP protein - Kennedy's yellow
PIR2:T04513	119.50	135.25	2.55	1334	DNA-binding protein R kappa
PIR2:J00532	119.50	129.91	2.88	556	probable membrane protein Y0
PIR2:S52865	119.00	137.51	2.59	915	hypothetical protein ZK945.1
PIR2:T21460	119.00	134.11	2.80	915	hypothetical protein ZK945.1
PIR2:T21460	119.00	132.79	2.88	915	clathrin assembly protein AP
PIR2:S53627	118.50	141.24	2.94	351	ORF5 protein - Ort virus (st
PIR2:B34768	118.50	134.53	2.56	713	neuregulin-3 [imported] - mo
PIR2:T44447	118.00	131.64	3.37	907	phosphoenolpyruvate carboxyl

seq_name: PIR2:H81782

seq_documentation_block:

adhesin MafB2 NMA2113 [imported] - Neisseria meningitidis (strain Z2491 serogroup A)

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: H81782

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo

; Halloway, S.; Jørgensen, K.; Leather, S.; Mouton, S.; Kung'u, K.; Quail, M.A.; Rajandre

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491

A:Reference number: AB1775; MIMD:20222556

A:Accession: H81782

seq_name: p1r2:B82028
seq_documentation_block:
adhesin NMA0324 [imported] - Neisseria meningitidis (strain Z2491 serogroup A)
C:Species: Neisseria meningitidis
C:date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: B82028
R:Parhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel, R.; Parhill, S.; Jørgensen, K.; Leather, S.; Moulé, S.; Mungall, K.; Quail, M.A.; Rajadream, Nature 404, 502-506, 2000
A:title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A8175; MUID:2022556
A:Accession: B82028
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1482 <PAR>
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83629.1; PID:g737908
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: mafB; NMA0324

alignment_scores:
Quality: 419.50 Length: 447
Ratio: 1.665 Gaps: 18
Percent Similarity: 56.376 Percent Identity: 31.320

alignment_block:
US-09-303-518D-465 x B82028 ..

Align seg 1/1 to: B82028 from: 1 to: 482

```
31 CTGTCCATCTAGCGAGTGTGCTGCCGATGATGACACGCGCTCAGATT 80
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
16 lIleSerLeuEnglnIleProIle.....SerHisAlaasnGlyLe 29
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
81 GGCAACGAGTTCTTTATCCGGCAGGTTTCGACCGTCAGCATTTGCAAC 130
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
29 u.....AspIaIArgLeuAlaArgAspMetGlnAlaIleHisIstGlu 44
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
131 CCGACGGGAATACCACTATTGCGC...AGCAGGGGGGAACCTTGGCGAG 177
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
44 roGlyGlyLysTyrHisLeuPheGlyAsnAlaIArgGlySerValLysAsn 60
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
178 CGCAGCGCATATCGGATGGGAAACATACAAAGCCATCATGTTGGGCAA 227
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
61 ArgVal.....TyrAlaValGlnThrPheAspAlaThrAlaValGlyPr 75
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
228 CTTGTC.....ATCCAGCAGGGGCGCATTAAGAAATATCGCT 268
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
75 oIleLeuProIleThrHisGlnArgThrGlyPheGlnGlyIleIleGly 92
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
269 ACATGTCGCGTTTCGATCAGGGGACAGATGCTATCCCGCTTGAC 318
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
92 yrgIuThrHisPheSerGlyHisGlyValHisSerProPheAsp 108
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
319 AACCATCCGCATATCGGATTCGATGAGCGGCTGATCCCGTTGACG 368
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
109 AsnHisSerLysSerThrSerAspPheSerGlyValAlaIspGly 125
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
369 ATTCCAGCTTACCGCATTCATCGGATGAGGATAGACACCATCCGCGC 418
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
125 yPheThrValTyrGlnLeuHisArgThrGlySerGlnIleHisProGlu 142
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
419 AGCGCTTACGCGGACAGGGGCGGCTATCCCGCTCCCAAGGCGCG 468
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
142 spGIyTyrAspGlyProGlnGlySerAspTyrProProGlyGlyAla 158
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
469 AGGATATATACAGCTACGACATTAAGGCGTGGCCCAAAATATCCGCT 518
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
159 ArgAspIleTyrSerTyrTyrValLysGlyThrSerThrLysThrLys 175
```

```
519 CAACCTGACCGCAACCGACGACCGCAACAGCGCTGTGACGCTTCC 568
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
175 rAsnIleValIProArgAlaProPheSerAspArgThrProLysGluAsn 192
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
569 ACAATACCGGTAGTATCTGACCGCAAGATAGCGGACGATTCAAACG 618
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
192 lAgIyAlaIleSerGlyPhePheSerArgAlaIspGlyAlaGlyLysLeu 208
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
619 GGCACCGCTACAGCCCGGACGTCGACAGATGGGCATGGCCGCGACG 668
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
209 lIeThrPheSerAspProAsnLysAsnThrPheAlaIspMetAspAs 225
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
669 TTTCACGCGCACTGCAAGATATGTCATAAC.....ATCATCGCGCGG 712
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
225 pIleArgGlyIleValGlnGlyAlaValAsnProPheLeuMetGlyPhe 242
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
713 CAGGAGAAATTTGCGCGCA.....GGCGATCCGCTGACGGGTAAAC 756
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
242 lGlnIyAlaGlyIleGlyAlaIleThrAspSerAlaValSerProValThr 258
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
757 GAAGGCTCAACATTCGTTGTCACGCG.....TTGGCTGTGCT 797
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
259 AspThrAlaIleGlnGlnThrLeuGlnGlyIleAsnHisLeuGlyAsn 275
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
798 TTCCACCGAAACAGATGGCGCGCATCAGCATTTGGCAGATATGCGG. 846
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
275 uSerProGlnAlaGlnIleuAlaIleAlaThrAlaLeuGlnAspSerAla 292
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
847 ..CAACTCAAGACTATGCCGCGACGACGATCCGGATTTGGCACTCCA 894
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
292 heAlaValLysAspGlyIleAsnSerAla...ArgGlnThrAlaIspAla 307
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
895 AACCCCAAT.....GCCGCAAGAGCATAGAACCGTCGCAAAAT 935
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
308 HisProAsnIleThrAlaThrAlaGlnThrAlaIleuAlaValAlaGlnAl 324
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
936 CTTTACGCGCATCCTCC..... 954
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
324 aaIaThrThrValITrpgIyLysLysValGlnIleuAsnProThrLysT 341
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
955 .....GTCAAGGATTTGA.....GCTTTCGGGGGAAA 984
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
341 rPAspTrpAlaLysAsnThrGlyTyrLysThrProAlaValArg..... 355
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
985 TACGCTTGGCGCGCATCAGGACATCTCTCAAGCGGTCGACATGGG 1034
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
356 .....ThmethIsthrLeuAspGlyGlnMetAlaGln 366
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1035 CGAGATCGCATTTGCCGAAAGGAAATCCGCGCTCAGCGCAATTTGCCG 1084
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
366 yGlyAsnArgProPro.....LysSerIleThrSerAsnSerLysAla 381
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1085 ATGCG..... 1089
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
381 sPAlaSerThrGlnProSerLeuGlnAlaGlnIleuIleGlyGlnIle 397
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1090 .....GCATACGCAATATC.....CCGTCCCTTACCA 1118
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
398 SerSerGlyHisAlaTyrAsnLysHisValIleArgGlnGlnIleuPhe 414
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1119 TTCCGCAATATCTGTTCAACTTGAGCAGCGCTTACGCAAGAAACA 1168
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
414 rAspLeuAsnIleAsnSerProAlaAspPheAlaArgHisIleGlnAsn 431
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1169 TCACCTCTCAACCGTCCCGCTCAAAACGAAAGAAATATG 1209
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
431 leValSerHis.....ProThrAsnMetLysGlnIleu 441
```

seq_name: p1r2:T4814

seq_documentation_block:
hypotheetical protein 156.220 [imported] - Neurospora crassa
C:Species: Neurospora crassa

C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000

C/Accession: T48814

R/Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, April 2000

A/Reference number: 224541

A/Accession: T48814

A/Status: Preliminary

A/Molecule type: DNA

A/Residues: 1-1952 <SCD>

A/Cross-references: EMBL:AL353822; GSPDB:GN00112; NCSP:1556.220

A/Experimental source: cosmid contig 15E6, strain 74

A/Genetics:

A/Gene: NCSP:1556.220

A/Map position: 2

A/Introns: 281/3

alignment_scores:

Quality: 170.00

Ratio: 0.742

Percent Similarity: 44.553

Length: 514

Gaps: 24

Percent Identity: 23.735

Alignment block:

US-09-303-518d-465 x T48814 ..

Align seg 1/1 to: T48814 from: 1 to: 1952

```
125 TCGAACCCGACGGGAGAAATACACATTCGAGCAGAGGGGAGAACTTGGC 174
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1221 AATThPrAlaAlaProValHISAlaThPrAlaThPrAlaLeuR 1237
175 GAGCGCAGCGGTCATATCGGATTGGGAACATCAAAAGCCATCAGTTGGG 224
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1237 oAla.....AsnHISHisThrMetAlaProA 1246
225 CAACCTGTCATCCAGCAGCGGCGCATTAAGAAATATCGGTACATTGG 274
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1246 lATHAlaProSerAlaAlaGlnAlaThPrAlaThPrAlaThPrAlaGln 1262
275 .....TCGCGTTTCCGATCAGGCGGCG 297
1263 ArgAlaAsnProProGlyGlyAsnProSerAlaValAlaProArgThrIleTh 1279
298 GAAGTCATCCCGCTTCGACAAACATGCTCAGATT.....C 335
1279 rAlaAlaAlaProAlaProThPrAlaProLeuProArgAlaAlaPro 1296
336 CGATTCTGATGAAGCGGATAGTCCGTTGACGATTCAGCCTTACCGCA 385
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1296 roIleSerIyHISProIleAlaProSerThrValAsnAsnThrAsnAla 1312
386 TCCATTGGGAGGATACGAACACCATCCGCGAGCGGTATGACGGGC.. 433
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1313 AlameGlyAlaArgPro.....ValProAlaAlaValAlaProGlyPr 1326
434 .....CACAGGGCGGCGCTATCCCGCTC..... 457
1326 oGlnIleGlyValHISSerGlyAlaAlaIleGlyAlaThPrAlaProValS 1343
458 .....CCAAAGCGCGAGGATATATACAGCTACGACATTAAGA 496
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1343 eArgSerProProValThrHISGlyValAlaThPrAlaAla..... 1356
497 GCGTTGCCCAAAATATCCGCTCAACCTGACGAGCAACCGCACCGACCGGA 546
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1357 .....ProThrThrThrProArgSe 1363
547 CAACGCTTGTGACCGTT.....T 566
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1363 rAsnLeuAlaAlaSerValProAlaGlySerMetAlaGlnIleThrHIS 1380
567 CCACATATACCGGTAGTATGTCAGCGAAGAGTAGGCGAGGATTCGAAC 616
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
1380 eArgValAlaProAla.....GlySerGlyAsnAlaSerIleSerArg 1393
617 GCGCCACCGCATACAGCCCGAGCTGACAGATGGGCAATGCCCGCA 666
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1394 AlaProSerThrAlaAlaProThPrAlaProThSerValProProVal 1410
667 GCTTTCACGGCAGCTGACATATCGTCAAAACATCATCGCGCGGCGAG 716
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1410 lSerSerThrValArgProMetSerIyValProThrGly.....Prop 1425
717 AGAAATGTTCGGCGCAGGGATGCCGTGCAGAGGTATACGAAAGCTCAA 766
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1425 roGlyLeuThrAla..... 1429
767 ACATTGCTGTATGACAGCGTTGGTGTCTGCTTCCACCGAAACAGATG 816
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1430 .....ProThrAlaSerSe 1435
817 GCGCGCATACAGATTGTGCAGATATGGCGCAACTCAAGACTATGCCG 865
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1435 rGlyAlaSerAlaGlyTyrAlaArgProAlaAsnAlaSerThrMetProA 1452
866 .....CAGCAGCCATCCGCGATTGGCGATGCCGCAAA 895
1452 lAProThrSerGlyGlnAlaAlaMetThrSerAlaValProGlnSer... 1467
896 ACCCAATGCCGACAAAGCATAGAACCGCTGACGAAATCTTTACGGCA 945
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1468 ValProSerPro.....ArgProSerSerLeuThrIleGly 1480
946 GTCATCCCGCTCAAGAGGATTGGAGCTGTTCGGGCAAAATACGGCTGGG 995
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1480 n.....MetGlyValProAlaIleAla 1488
996 CGGATCAGCGGACATCTGTCAAGCGGTGCGAGATGGCGAGATGCAT 1045
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1488 lAlaSerThrSerArgProAlaSerGly..... 1497
1046 TCGCGAAAGGAATCGCGCTGACGACAAATTTCCGATCGGCGCAT.. 1093
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1498 .....ValTyrAsnProProAlaSerSerLeuAlaProSerThrHISly 1512
1094 .....ACGCCAATACCCGCTTCACATTCGCCGAATATCCGTTCAAA 1139
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1512 sSerMetProSerAlaValProThrThrAlaSerGlyAlaValSerSeT 1529
1140 CTTCGACGACGCTTACGCGCAAGAAACATCACTCCATCAACCGTGS... 1185
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1529 hrValSerSerLeuAlaAlaThrProLeuProProSerAlaProArgTyr 1545
1186 CCGCGCTCAACGGAAGATGTGAAACTGGCAMAACCAAGCCACCGCAA 1235
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1546 GlyProSerSerAsnAlaThrProThrThrProAlaThrAlaIleProAr 1562
1236 GACCAAAAGTCCGTTGACGTAAG.....GGGTTTC 1267
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1562 gAlaProLeuProAlaAlaSerSerValSerAlaProValThrGlyGlnP 1579
1268 CGAATTTTGAAGAAAGACGTAATAACGATACGAGTAATATACCGCTGA 1317
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1579 roSerPhSerAlaProAlaSerValProThrProProThrSerAlaAla 1595
1318 CCACAA...GTGAATCTATTAAGTGAACCGCTTTAACTCT..... 1356
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1596 serGlnGlyAlaGlnProLeuSerGlnProVal...GlyProAlaProAl 1611
1357 .....AAAGTCTGTGCGATCGGCTCATCTGTGCTATTAACGCGCAGA 1402
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1611 alIleSerSerIleGlyAlaThrProAlaAlaSerIleProSerSeA 1628
1403 TTCAATATCCCAAAATTACCAAGCAAGATGATATATATATATATATAT 1449
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
1628 lAProAlaAlaAlaAlaPro.....ValThrTyrProValPro 1640
```

1450 CCTAAATTAATCTCTCTTACAGCCGCTACCAAGGACCT 1491
 1641 GlnGlnAlaSerAlaAlaAlaAlaAlaArgLeuProValThrPro 1654
 seq_name: p1r2:T45025

seq_documentation_block:
 mucin MUC5B, tracheobronchial [imported] - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T45025
 R:Desseyn, J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
 A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
 A:Reference number: 222899; MUID:97166151
 A:Accession: T45025
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
 A:Experimental source: placenta
 C:Genetics:
 A:Gene: MUC5B

alignment_scores:
 Quality: 159.50 Length: 430
 Ratio: 0.794 Gaps: 20
 Percent Similarity: 46.744 Percent Identity: 23.721

alignment_block:
 US-09-303-518D-465 x T45025 ..

Align seg 1/1 to: T45025 from: 1 to: 3570

12 CCGCAAAATATCCCTATCTGCTCCACTAGCGAGTGGCCCGGATGC 61
 1197 ProSerSerThrProGlyThrAlaHisThrLysValProThrThr 3213
 62 ATGCAC.....A 68
 3213 ThrThrGlyPheThrAlaThrProSerSerProGlyThrAlaLeu 3230
 69 CGCCTCAGATTGGCAAGCATCTTTATCCGAGGTTCTCCAGCGTC 118
 3230 hrProProValThrPheSerThrThrThrProThrThrThrPro 3246
 119 AGCATTGGAAACCGAGCGGAATACCACTATTCCGAGCGAGGAA 168
 3247 ThrThrSerGlySerThr.....ValThrProSerSerIleProGly 3261
 169 CTGGCCGAGCGAGCGTCATATCGGATTGGAAACATACAAAGCCATC 217
 3261 rThHisThrAlaArgValLeuThrThrThrThrValAlaThrG 3278
 218 ..AGTTGGCAACCTGTTCAATCAGAGCGGCGCATTA.....AAG 256
 3278 LysSerMetAlaThrProSerSerSerThrGlnThrSerGlyThrPro 3294
 257 GAAATATCGGTCATATGTCGCTTTCCATCAGGCGAGCAAGTCCAT 306
 3295 SerLeuThrThrThrAlaThrThrIleThrAlaThrGlySerThr 3311
 307 TCCCCCTTCGACA.....ACATGCTCAACATCCGATCTG.....A 344
 3311 nProSerSerThrProGlyThrThrProIleProProValLeuThrSer 3328
 345 TGAAGCCGTAAGTCCGTTGACGAGTCAACCTTACCGCATCATGGG 394
 3328 etAlaThrThrProAlaAlaThrSerSerLysAlaThrSerSerSer 3344
 395 ACGGATACG...AACACATCCGCGCAGCGCATATGAGGGCCACAGGGC 441

3345 ProArgThrAlaThrThrLeuProValLeuThrSerThrAlaThrLys 3361
 442 GGGCGCT.....ATCCGCTCCCAAGCGCGAGGATATATACAG 482
 3361 rThrAlaThrSerPheThrThrProIleProSerSerThr...LeuThr 3377
 483 CTACGACATAAAAAGCG.....TTGCCAAATATCCGCTCAGC 523
 3377 hrThrPheValProAlaGlnThrThrThrProMetSerThrMetSer 3393
 524 TGACGACACACCGCAGACCGGAG...AAGGCTTGCGACCGTTCCAG 570
 3394 IleHisThrSerSerThrProGlyThrThrHisThrSerThrVal 3410
 571 AATACCGTAGTATGCTGACGACAGAGAGAGCGGATTCAAAGCGC 620
 3410 rThrThrAlaThrMetThrArg.....AlaThrHisSerThrAla 3424
 621 CACCCGATACAGCCCGAGCTGACAGATCGGCGCAATGCCCGAAGCT 670
 3424 hrProSerSerThrLeuGlyThrThr.....ArgLe 3434
 671 TCACGCGCATGCAAGATATGTCMAAAACATCATCGGCGCGGAGAA 720
 3435 LeuThrGluLeuThrThrThrAlaThrThrThrAlaAlaThrGly 3451
 721 ATGTGCGCGCAGCGAGATCCGTCAGGGTATTAAGCAAGGCTCAACAT 770
 3451 rAlaThrLeuSerSerThrPro..... 3458
 771 TGCCTTATGACGCGCTTGCTGCTCTTCCACCGAAACAGATGCGCGC 820
 3459GlyThrThrThrPheLeuThrGluPro..... 3467
 821 GCATCAAGATTTGGAGATATGCGCAGACACCAAGACTATGCCGCGCA 870
 3468 ...SerThrIleAlaThrValMetValProThrGlySerThrAlaThr 3483
 871 GCCATCCGCGATTTGGCATGCCAAACCCCAATGCCGACAGGCAATGA 920
 3483 rSerSerThrLeuGlyThrAlaHisThr..... 3492
 921 AGCCGTACGAATATCTTACGGAGCATCCCGTCACAAAGGATTTGAG 970
 3493 ...ProLysValAlaThrAlaMetAlaThrMetPro..... 3503
 971 CTGTTGGGGAATAATACGGCTTGGCGGATCAGGACATCCTGTCAAG 1020
 3504ThrAlaThrAlaSerThr..... 3509
 1021 CGGTCCGAGATGGGCGAGATCGCATTCGGAAGGAATCCGCGTCAG 1070
 3510ValProSer 3513
 1071 CGACAAATTTGCCGATCGGATACGCAATATACCGCTCCCTTACCATT 1120
 3513 erSerThrValGlyThrThrArgThrProAlaValLeuProSerSer 3529
 1121 CCGGAATATCGCTTCAACTGGAGCAGC.....GTTACGGACA 1161
 3530 ProThrPheSerValSerThrValSerSerValLeuThrThrLeu 3546
 1162 GAACATCAGCTCTCAAC.....CCGTGC 1186
 3546 gProThrGlyPheProSerSerHisPheSerThrProCys 3559

seq_name: p1r2:T30826
 seq_documentation_block:
 nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
 N:Alternate names: alpha-MNC protein
 C:Species: Mus musculus (house mouse)

C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C/Accession: T30826
R/Y/Olov, W.V., St-Arnaud, R
Genes Dev. 10, 1763-1772, 1996
A>Title: Differential splicing in of a proline-rich exon converts alphaNAC into a muscle
A/Reference number: z20889; MUID:96312450
A/Accession: T30826
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2187 <YOT>
A/Cross-references: EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AMB18732.1
C/Genetics:
A:Gene: Naca
A:Map position: 10
A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A/Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding actly
C/Keywords: alternative splicing; DNA binding; transcription factor

alignment_scores:

Quality: 156.50 Length: 567
Ratio: 0.660 Gaps: 24
Percent Similarity: 41.799 Percent Identity: 20.988

alignment_block:

US-09-303-518D-465 x T30826 ..

Align seg 1/1 to: T30826 from: 1 to: 2187

```
167 AACATGCGAGCGAGCGGTCATATCGAATGGGAACATTAACAGCCAT 216
      ::::::::::::::: ||| ::::| ::::| ::::|
890 SerIleProlySValThrSerProSerProGlnThrProlySValPro 906
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
217 CAGT.....TGGCAACCTGTCATCCAGCAGCGCCATTAAGGAA 260
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
906 IserIeuylSgIyAlaProAlaMetThrSerIyls...AlaThrGlu 922
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
261 TATGCGCTACATGTCGCTTCCTTCGATCAGCGGCACAGATCCATTC 310
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
922 IeAlaIleSerIylsAspIalSerProSerGlnPheProlySgIValPro 938
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
311 CCTTCGACACCATGCTCACATTCGATTCGTATGAAGCCGGTAGTCC 360
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
939 .....LeuIeuGlnIleValIleProThrProThrSerProThrSerp 952
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
361 GTTGCAGATTCA.....GCCTTACCGCA..... 385
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
952 oValSerAspThrLeuSerGlyAlaLeuThrSerProProProlySgIYP 969
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
386 .....TCCATTGGACGATACAGAACACCATCCG.....CCGACG 421
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
969 roProAlaThrLeuAlaGluThrProThrTyProlySerpProlyS 985
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
422 GCTATGACGGCGACAGCGCGGCTATCCGCTCCCAAGAGCGGAGG 471
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
986 ProAlaIleSerIylsThrProAlaThrProSerProGluGly.... 1000
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
472 GATATATACAGCTACGACATAAAGCGCTGCCAAATATCCGCTCA 521
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1001 .....ValThrAlaValPro...LeuGluIleProProGlySerIylsYA 1015
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
522 CCTGACCGAACAACCGACGACGAGGCTGTGCGACGCTTTCACCA 571
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1015 1a...ProlySThrAlaIleProlySgIuIleSerIalThrSerSer... 1029
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
572 ATACCGGTATGATGCTGACGCAAGAGTAGCGACGATTTCAACGCGCC 621
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1030 .....SerIylsArgAl 1033
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
622 ACCGATACAGCC.....CCGAGCTGACAGATCGGGCA 656
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1033 apProlySThrAlaValSerIylsGluIleProSerIylsGlyValThrAlav 1050
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
```

```
657 TGGCCGCCAGCTTTCACAGCGCACATGATATGCTCAAAACATCATCG 706
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1050 alProleuGluIleSerLeuProleuSgIuThrSerIylsSerAlaThr 1066
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
707 GCGGCGACGAGGAATTTGCGCGCAGCGCATGCGTCAGGATTAAGC 756
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1067 ProGlyIuIylsSerAlaSerSerProlySArgSerProlySThrAlaG 1083
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
757 GAAGGCTCAACA.....TGGCTGTATGACGCGCTGGGTC 794
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1083 yProlySgIuThrProProGlyIylValThrAlaValProProGluIles 1100
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
795 GCTTTCACCGCAAAACAAGATGCGCGCATCACGATTTGGCAGATATGG 844
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1100 erIeuProProlySgIuThrProGlnAlaThrProAsnGluSerLeu 1116
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
845 .....CGCACTCAAAAGCT..... 859
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1117 AlaAlaSerSerGlnIylsArgSerProlySThrSerValProlySgIuTh 1133
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
860 .....ATGCCGACGACCCAT 875
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1133 rProProGlyIylValThrAlaMetProleuGluIleProSerAlaProG 1150
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
876 CCGCATTTGGCGATCCCAAAACCCA..... 901
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1150 IuIylsAlaProlySThrAlaValProlySgIuIleProThrProGluasp 1166
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
902 .....ATGCCGACAGGCTAGAA 921
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1167 AlaValThrIleLeuAlaGlySerProleuSerProlySValaSerIly 1183
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
922 GCCGTCAGCAATATCTTACGGCAGTATCCCGCAAGGAAATCCCGTCAGC 971
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1183 sThrAlaIleProlySgIuIleProAlaThrProSerValGlyVal.... 1198
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
972 TGTTCGGGGAATAACGCTTGGCGCATCACGCGCATCTGTCAAGC 1021
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1199 .....IleAlaValSerGlyIuIleSerProSer 1208
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1022 GCTCGAGATGGCGAGATGCGATTCGCCGAAGGAAATCCCGTCAGC 1071
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1209 ProlySThrSerIylsThrAlaIleProlySgIuAsn.....SerAl 1223
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1072 GAAATTTTCCGATGCGCATACGCCAATACCCGCTTACCATTC 1121
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1223 aThrIleProProlySArgSerProlySThrAlaIleProlySgIuThrP 1240
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1122 CCGAAATATCCGTTCAAACTTGGACGACGCTTACGCAAG...AAAACA 1168
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1240 roAlaThrSerSerGluGlyValThrAlaValProSerGluIleSerPro 1256
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1169 TCACCTCTCTCAACCGTCCG..... 1188
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1257 SerProProThrProAlaSerIylsGlyValProValThrLeuThrPro 1273
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1188 ..... 1188
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1273 sGlyAlaProAsnAlaLeuAlaGluSerProAlaSerProlySValP 1290
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1189 .....CCGTCAAACGGAAG 1203
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1290 rolySThrAlaIleProGluGluThrSerThrThrProSerProGlnIyls 1306
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1204 AATGTGAACAGCGCA.....AACAA 1223
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1307 IleProlySValAlaGlyProlySgIuAlaSerAlaThrProProSerly 1323
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1224 ACGCCACCCGAAGCAAAAGTCCGTTGACGGTAAGGGTTCCGAT 1273
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1323 slySThrProlySThrAlaValPro..... 1331
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
1274 TTGAAAAAGAGCTAAATATGATACGAGAAATTAATACCGCTGACCA... 1320
      ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
```

1332LySGLuThrSerAlaProSerGluGlyValThrlAlaValProLeu 1346
1321CAAGTAAATCCATAGA 1337
1347 GuileProProSerProAlaGlyAlaProLysThrAlaAlaProLysG 1363
1338 TGACCCGCTTTTAATCTAAAGGTTCTGCGAGTGGCTCATCTTGGT 1387
1363 uThrProAlaProSerProGluGlyAlaThr..... 1373
1388 CTATACGCCCAATTCATATACGAAATATCCAGCAGGATGAAATC 1437
1374ThrlAlaProValGlnleProProSerProAlaGlySerLys 1388
1438 AGATATATCCACCTAAATATACCTCCATGACGAGCTATCCAAAGG 1487
1389 Lys.....AlaGlySerLysGluThrProThrProSerProGluG 1403
1488 A 1488
1403 Y 1403

seq_name: p1r2:A43932

seq_documentation_block:
mucin 2 precursor, intestinal - human (fragments)
N/Alternate names: mucin SMUC-41
C/Species: Homo sapiens (man)
C/Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #ext_change 05-Nov-1999
C/Accession: A49963; A45106; B45106; A43932; B33532; A61257; P00328; P00329
R/Gum JR., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A>Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A/Reference number: A49963; MUID:94132002
A/Accession: A49963
A/Molecule type: mRNA
A/Residues: 1639 <GU1>
A/Cross-references: GB:I21998
R/Gum JR., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A>Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr
A/Reference number: A45106; MUID:93016075
A/Accession: A45106
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 626-1895 <GU2>
A/Cross-references: GB:M94131; NID:9186395; PIDN:AAA59163.1; PID:9186396
A/Note: sequence extracted from NCBI backbone (NCBIP:110706)
A/Accession: B45106
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 2037-3020 <GU3>
A/Cross-references: GB:M94132; NID:9186397; PIDN:AAA59164.1; PID:9186398
A/Experimental source: colon
A/Note: sequence extracted from NCBI backbone (NCBIP:116698)
R/Toribara, N.W.; Gum JR., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.
J. Clin. Invest. 88, 1005-1013, 1991
A>Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp
A/Reference number: A43932; MUID:91358717
A/Accession: A43932
A/Molecule type: DNA
A/Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A/Cross-references: GB:M74027; NID:9188863; PIDN:AAA59875.1; PID:9188864
A/Note: sequence inconsistent with the nucleotide translation
A/Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)
R/Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A>Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden
A/Reference number: A33532; MUID:89197956
A/Accession: B33532
A/Molecule type: mRNA
A/Residues: 1916-2193 <GU4>

A/Cross-references: GB:M22405; NID:9188873; PIDN:AAA36334.1; PID:9188874
A/Experimental source: intestine
R/Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
J. Clin. Invest. 87, 77-82, 1991
A>Title: Human bronchus and intestine express the same mucin gene.
A/Reference number: A61257; MUID:91086481
A/Accession: A61257
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
A/Experimental source: bronchus
R/Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCoil, D.; Wang, D.; Jones, C.; Forstn
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A>Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the
A/Reference number: P00328; MUID:92198477
A/Accession: P00328
A/Molecule type: mRNA
A/Residues: 2328-2468 <XUG>
A/Cross-references: GB:M86523
A/Experimental source: small intestine
A/Accession: P00329
A/Molecule type: protein
A/Residues: 2328-2342, 'K', 2344-2354 <XUG1>
A/Genetics:
A/Gene: GDB:MUC2
A/Cross-references: GDB:120203; OMIM:158370
A/Map position: 11p15.5-11p15.5
C/Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; v
C/Keywords: glycoprotein; intestine; tandem repeat
F:2766-2834/Domain: von Willebrand factor type C repeat homology <WVC>

alignment_scores:

Quality:	153.50	Length:	315
Ratio:	1.104	Gaps:	13
Percent Similarity:	44.127	Percent Identity:	24.444

alignment_block:

US-09-303-518D-465 x A43932 ..

Align seg 1/1 to: A43932 from: 1 to: 3020

269	ACATTGCGCTTTTCGATCAGCGGAGCAAGTCCATCCCTTCGAC	318
1520	ThrThrThrProSerProThrThrThrThrThrThrProProProth	1536
319	AACCATGCTCATTCGATCTGATGAAGCGGAGTCCGTTGACGG	368
1536	rThrThrProSer...ProProThrThrThrProThrProProProthS	1552
369	ATTCAGCTTTTACCGCATTCATGGAGCGATATGCAACATCCGCGC	418
1552	erThr.....ThrThrLeuProProThrThrThrThrProProPro	1566
419	ACGGCTATAGAGGGCCACAGGGGGCGCTATCCGCTCCAAAGCGCG	468
1567	ThrThrThrThrThrProProProProThrThrThrProSerPro.....	1580
469	AGGAGATATATACAGCTACGATGCAATAAAGCGCTGCCCAAAATACCGCT	518
1581ProThrThrThrThrThrProProProProThrThrThrT	1594
519	CAACCTGACCGACACCGACGACGACGAGGCTGTCGACGCTTCC	568
1594	hThrProProProThrThrThrThrProSerProProThrThrThrThr	1610
569	ACAAATCCGGTGTATGCTGACGCAAGAGTGGCGAGGATCAACGC	618
1611	ThrProPro.....ProThrThrThrProSe	1619
619	GCACCGATACGACCGCGAGCTGGACAGATGGGCAATGCGCGGAGC	668
1619	rProProThrThrThrPro.....lleThrProProThrs	1631


```

732 AGCGATGCGCCGAGGTATTAAGCGAAGCGTCAACATTGCTGTATGC 781
    ||| |||||
284 ArgAlaArgArgAlaGly..... 289
782 AGCGTTGGCTTCCTTCCACGGAACAGAT..... GCGCGCAT 824
    ||| |||||
290 .....GlyGlyGlyLeuProArgArgGlyAspArgValArgAlaHis 304
825 .....CAACGATTGGCAGATATGGCGCACTCAAGACTATGCGCGAG 869
    ||| |||||
304 rgrProArgProArgGlyArgGlyAlaAla.. ArgLeuAspArgAla 319
870 AGCGATCGCGATTGGCGAGTCCAAACCCCATGCCGCAAGGCATAG 919
    ||| |||||
320 GlyGlyAlaGlyAlaArgProAlaArgProArg... ArgLeuArg..... 333
920 AAGCGTCAGCAATATCTTTACGCGAGTCATCCCGCTCAAGGGATGGA 969
    ||| |||||
334 .....HisProArgProAlaAspLeu 341
970 GCTGTGCGGGAATAACGGCTTGCGCG... CATCAGCGCATCTCT 1016
    ||| |||||
341 rGHisGlyGlyProGlnGlyAlaAlaArgLeuAspHisProGlnGlyLeu 357
1017 CAAGCGTCGCGAGATGGCGCGATCGCATCCGCAAGGGAATCGCGCG 1066
    ||| |||||
358 GluGlyAlaAlaGlySerArg..... 364
1067 TCAGGACAAATTTCGCGATGCGGCAATACGCCAATACCCGCTTAC 1116
    ||| |||||
364 ..... 364
1117 CATTCGCCGAATATCCGTTCAACTTGAGAGCGGTTACGCAAGAAAA 1166
    ||| |||||
365 .....ArgHisProHisArgLeuArgAlaArgLeuHisGlnGlyArg 378
1167 CATCACTCTCTCAACCGTCGCGCGCTCAACGGAAGATGTGAACCTGG 1216
    ||| |||||
379 GlyAspLeuLeuArgArgProArgArgAspArg..... LeuG1 391.
1217 CAACAAACGCGCACCGCAAGACCAAGTGC 1247
    ||| |||||
391 YArgArgGlyProArgGlnGlyAla 401

```

seq_name: p1r2:S35047

seq_documentation_block:

```

mucin JUL7 - human
C:Species: Homo sapiens (man)
C:date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 08-Oct-1999
C:Accession: S35047
R:Dufose, J.; Porchet, N.; Audle, J.P.; Guyonnet Duperrat, V.; Laine, A.; van-Seuningen,
Biochem. J. 293, 329-337, 1993
A>Title: Degenrate 87-base-pair tandem repeats create hydrophilic/hydrophobic alternati
A:Reference number: S35047; MUID:9343838
A:Accession: S35047
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-543 <DUP>
A:Cross-references: EMBL:X74370; NID:9407081; PUID:CA52408.1; PID:9407082
A>Note: the authors translated the codon AAA for residue 63 as Gln and CGG for residue 3
A>Note: the amino acid sequence from fig. 2 is inconsistent with the nucleotide sequence

```

```

alignment_scores:
  Quality: 149.50      Length: 395
  Ratio: 0.808        Gaps: 16
  Percent Similarity: 46.835      Percent Identity: 22.025

```

alignment_block:

US-09-303-518d-465 x S35047 ..

Align seg 1/1 to: S35047 from: 1 to: 543

```

128 AACCCGAGGGGAAATACACCTATTTCGACAGGAGGGAATTCGCGAG 177
    ||| |||||
44 SerProLeuAspSerThrAlaThrProSerThrProGlyThrAlaPro 60
178 GCGACGCGTCATATCGATTGGGAACATCAACAGCCATCTGGCGCA 227
    ||| |||||
60 ProGlyValLeuThrSerProAlaThrThrProThrAlaThrSerSer 77
228 CCGTTCATCCACGACGCGGCATTAAGGAATATAGCGGTACATTCGC 277
    ||| |||||
77 yAlaThrSerSerSerPro..... ArgThrAlaThrThrLeuPro 91
278 GCTTTCGATCAGGCGACGAATGCAATTCCTCCCTTGACACCATGCC 327
    ||| |||||
92 ValLeuThrSerThrAlaThrLysSerThrAlaThrSerValThr.... 106
328 TCACATTCGATTCGATGAAAGCCGTAGTCCCGTTCGACGATTCAGCT 377
    ||| |||||
107 .....ProIleProSerSerThrLeuGlyT 115
378 TTACCGCATCCATTGGGAGCGATACGACACCATCCCGCGAGGCTATG 427
    ||| |||||
115 hrThrGlyThrLeuProGlnGlnThrThrThrProValAlaThrMetSer 131
428 ACGGCGCACAGGCGCGCGCTATCCCGCTCCCAAGGCGCGAGATATA 477
    ||| |||||
132 ThrIleHisProSerSerThr.....ProGlnThrThrHisThrSe 145
478 TCACGTCAGACATTAAGAGCGTGGCCAAATATCCGCGCTACACCTGAC 527
    ||| |||||
145 rThrValLeuThrThrLysAla..... 152
528 CGACACCGGACGACCGGACAAAGCGCTTCGACGCGTTTCACAAATACG 577
    ||| |||||
153 ..ThrThrThrArgAlaThrSerSerThrSerThrProSerSerThrPro 168
578 ...GTAGTATGCTGACGCAAGAGTAGCGGAGATTCAAAGCGCGCAC 624
    ||| |||||
169 GlyThrThrThrPrlLeuThrGlnLeuThrThrAlaAlaThrThrThrAl 185
625 CGATACAGCGCGCGAGTGGACAGATCGGCAATGCCCGCGAAG.....C 668
    ||| |||||
185 agGlyThrGlyProThrAlaThrProSerSerThrProGlyThrThrPrl 202
669 TTTCACGCGCATGTCAGATATGTCAAAACATCATCATCGCGCGCGAGAG 718
    ||| |||||
202 leuThrGlnLeuThrThrThrAlaThrThrThrAlaSerThrGlySer 218
719 AATTCGCGCGGCGAGCGATGCGGTGAGGCTTAAGCGAAGGCTCAAC 768
    ||| |||||
219 ThrAlaThrLeuSerSerThrPro..... 226
769 ATTGCTGTATGCAAGCGCTTGCGTTCCTTCCACGGAACAGATGGC 818
    ||| |||||
227 .....GlyThrThrThrPrlLeuThrGlnProSerThr..... 237
819 GCGCATCAAGATTGGCAGATATGGCGCAACTCAAAAGACTATGCGGAG 868
    ||| |||||
238 ..ThrAlaThrVal.....ThrValProThr 245
869 CAGCATTCGCGGATGGG.....CAGTCCAAACCCCAATGCCGCA 909
    ||| |||||
246 GlySerThrAlaThrAlaSerSerThrGlnAlaThrAlaGlyThrProH1 262
910 CAAGCATAGAACCGCTGACGAAATATCTTTACGAGCATCATCCCG... 955
    ||| |||||
262 sValSerThrThrAlaThrThrProThrValThrSerSerLysProLeu 279
956 .....TCAAGGAGTGGAGCTTCGCGGAAATATACGCTTGCGCGCA 1000
    ||| |||||
279 euleuSerProGlyLeuGlnLeuProSerAlaLeuArgSerThrAlaThr 295

```

```

1001 TCACGGAC.....ATCCTGTCAGCGTGC 1026
    :: |||
296 AsnThrHisSerTyrGlnLeuTyrSerHisProLeuLeuProGlyTh 312
    |||::|||
1027 CAGATGGCGCAGATGCCATTGCCGAAGAAATCCGCGTCAGCAGCAA 1076
    |||::|||
312 rThrThrProTyrArg.....LeuSerGlnThrThrThrProThrAlaThrM 327
    |||::|||
1077 TTTTGGCGATGCGCATACGCAATATACCGTCCCTTACATGCCGAA 1126
    :: |||
327 et.....SerThrAlaThrProSerSerThrProGlyThrValHisThr 341
    |||::|||
1127 ATATCCGTTCAAACTGTGAGCAGCGCTTACGCAAGAAATCA..... 1171
    |||::|||
342 SerThrValLeuThrAlaThrAlaThrThrThrGlyAlaProAlaLeuThr 358
    |||::|||
1172 ....CCTCTCAACGCGTGGCGCGTCAACGCAAGAAATGTAACTGCGC 1217
    |||::|||
358 pProProProProProProGlnGln..... 367
    |||::|||
1218 AAACAACCCACCCGAGACCAAGTGGCGTTTG 1252
    |||::|||
368 .....LeuThrLeuProLysCysArgLeu 375
    |||::|||
seq_name: p1r2:T22808

```

```

seq_documentation_block:
  hypothetical protein F56H9.1 - Caenorhabditis elegans
  C:Species: Caenorhabditis elegans
  C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
  C:Accession: T22808
  R:Burton, J.
  submitted to the EMBL Data Library, June 1996
  A:Reference number: Z19618
  A:Accession: T22808
  A:Status: preliminary; translated from GB/EMBL/DBJ
  A:Molecule type: DNA
  A:Residues: 1-770 <MIL>
  A:Cross-references: EMBL:Z74473; PIDN:CAA98949.1; GSFDB:GN00023; CESP:F56H9.1
  A:Experimental source: clone F56H9
  C:Genetics:
  A:Gene: CESP:F56H9.1
  A:Map position: 5
  A:Introns: 235/1; 262/2; 320/1; 367/2; 510/3; 654/1; 681/2

```

```

alignment_scores:
  Quality: 149.00      Length: 219
  Ratio: 1.263        Gaps: 9
  Percent Similarity: 53.881      Percent Identity: 25.571

```

```

alignment_block:
  US-09-303-518D-465 x T22808

```

```

Align seg 1/1 to: T22808 from: 1 to: 770

```

```

86 ACAGTTCTTTATCCGCGAGTTCTGACCGTACGATTCGAACCGCAG 135
    |||::|||
37 ThrValProValSerThrThrThrThrThrThrThrThrThrThr 53
    |||::|||
136 G.....GGAATATACCATTCGTCGCGAGGCGGGAAGTCCGAGCG 179
    |||::|||
53 rValProSerThrThrThrThrThrThrThrThrThrThrThrValProSer 70
    |||::|||
180 CAGCGGTATATCGATGATGGGAAACATACAAAGCCATGTTGGCAGAC 229
    |||::|||
70 hrThrThrSerSer.....Thr 75
    |||::|||
230 TGTTCATCCAGCAGCGCGCATTAAGAAATATCGCTACATTCGCGC 279
    |||::|||
76 ThrThrThrThrThrValProThrThrThrThrThrThrThrThrThr 92
    |||::|||
280 TTTTCCG.....ATCAGGCGCAGAGTGCATTCGCCCTT 314
    |||::|||

```

```

    |||
92 rValProProThrThrThrThrThrThrThrThrThrThrThrThrThr 109
    |||::|||
315 CGACACATGCTGCTTACATTCGATTCGATTCGATTCGATTCGATTCG 364
    |||::|||
109 hrThrThrSerThrThrThrThrThrThrThrThrThrThrThrThr 123
    |||::|||
365 ACGATTCAGCCTTACCGCATTCAT.....GGAC 396
    |||::|||
124 ThrSerSerThrThrThrThrThrThrThrThrThrThrThrThrThr 140
    |||::|||
397 GATATCGACACATCCCGCGACGCGTATGACGCGCAGCGCGCGCG 446
    |||::|||
140 rThrThrThrThrThrValProProThrThrThrThrThrThrThrThr 157
    |||::|||
447 CTATCCCGCTCCCAAGGCGGAGGATATATACGTACGATTAAG 496
    |||::|||
157 hrValPro...ProThrThrThrThrThrThrThrThrThrThrThr 170
    |||::|||
497 GCGTTCGCAAAATATCCGCTCAACCTGACCGACCAACCGCAGCGCG 546
    |||::|||
171 ..ValProAlaThrThrThrThrThrThrThrThrThrThrThrThr 185
    |||::|||
547 CAACGGCTGTGCGACGTTTCCACATACCGGATGATGCTGACGACAG 596
    |||::|||
186 .....ProThrThrSerThr..... 190
    |||::|||
597 AGTAGGCGAGGATTCAAACGCGCACCGCATACAGCCCGCGCTGACGA 646
    |||::|||
191 .....ThrThrThrThrThrValProProThrThrThrThrThrThr 205
    |||::|||
647 GATGCGGCAATGCGCGGAGCTTTCACGCGCAGCTGACATATCTCAAA 696
    |||::|||
206 ThrThrThrThrThrValProProThrThrThrThrThrThrThrThr 222
    |||::|||
697 AACATCA 703
    |||::|||
222 rThrThr 224
    |||::|||
seq_name: p1r2:T15316

```

```

seq_documentation_block:
  hypothetical protein B0302.1 - Caenorhabditis elegans
  C:Species: Caenorhabditis elegans
  C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
  C:Accession: T15316
  R:Du, Z.
  submitted to the EMBL Data Library, November 1995
  A:Description: The sequence of C. elegans cosmid B0302.
  A:Reference number: Z18328
  A:Accession: T15316
  A:Status: preliminary; translated from GB/EMBL/DBJ
  A:Molecule type: DNA
  A:Residues: 1-1223 <DUZ>
  A:Cross-references: EMBL:U01032; NID:g1086728; PID:g1086729; PIDN:AAA82367.1; CESP:B0
  A:Experimental source: strain Bristol N2
  C:Genetics:
  A:Gene: CESP:B0302.1
  A:Introns: 50/3; 106/3; 153/3; 180/3; 243/3; 399/2; 438/3; 482/1; 520/2; 545/1; 589/1

```

```

alignment_scores:
  Quality: 148.50      Length: 530
  Ratio: 0.640        Gaps: 23
  Percent Similarity: 43.774      Percent Identity: 22.453

```

```

alignment_block:
  US-09-303-518D-465 x T15316

```

```

Align seg 1/1 to: T15316 from: 1 to: 1223

```

```

12 CCGCAAAATATCCCTTATCTGTCCATATGCGAGTGTGCGCGATGC 61
    ||| ||||| ||| ||| ||||| ||

```

```

715 ProAlaAsnIleProCysIleValProThrProAlaProAlaProAl 731
62 ATGCACAGCGCTCAGATTGGCAACGATCTTTATCCGGCAGGTC 111
731 a.....HisPheSerGlnProValSerSerG 740
112 GACCGTCAGCAT..... 124
740 InArGValAlaGlnGlnGlnGlnAsnThrLeuGlnIleValAlaLeuAsnAr 756
125 .....TCGAACCGGAGGGAATACCACTATTC 153
757 GluLeuIleGlyAsnLeuAsnIleValArgProThrGlyThrThrAlaProPr 773
154 GGCAGCAGGGGGGAGCTGGCAGCGGTCATATCGATTGGGAAA 203
773 oSerAsnGlyPheAsnAlaProAlaAlaAspValAlaProValGlnGln 789
204 CATCAAGACCATCAGTTGGCAACCTGTTCATCCAGCAGG...CGGCCA 250
790 .....ArgProIleSerSerAlaSerIleProAlaLeuGlnProGlnPro 804
251 TTAAGGAATATCGGCTACATGTCGGCTTTCCGATCAGGGCAGAA 300
805 ILeGlnHisIleGlnIleProIleGlnProGlnGlnVal..... 817
301 GTCCATTCCTCCCTGACAAACCATGCTCAGATTCGATTCGATGAGC 350
818 ArgIleProProSerThrAlaProValGlnIleProValGlnValSerA 834
351 CG.....GTAGTCCCGTTGACGGATTCAGCCTTTACCGCA 385
834 LaProThrHisSerAsnValAlaProThrThrSerGlnAlaSerAla 850
386 TCCATTGGGAGGATACGACACCATCCCGCG... 418
851 AspAla.....ArgAsnProLeuProProLysThrSerProProVa 864
419 .....ACGGCTATGACGGGCCACAGGGCGCGGCTATCC 452
864 IserAsnThrProIleThrValAlaProValHisAlaAlaProThrThr 881
453 CGCTCCCAAGCGCGAGGATATAT.....ACAGTACGACATAA 493
881 eAlaIaProSerThrSerValValThrArgArgProThrSerThrAla 897
494 AAG..... 496
898 GlnMetSerAspGluGlnArgSerArgIleAlaMetAspIleSerSe 914
497 GCGTTGGCCCAAAATATCCGCC.....TCAACCTGACCGCAGACACG 536
914 rAlaIleProAlaProSerAlaIleLeuIleValGlySerAsnSerThr 931
537 CAGCAGCGGCAAGCGCTGTGACCGTTTCCACATACCGGTAGTATGC 586
931 eIleuProSerAlaAlaValSerThrAlaSerSerValPro..... 944
587 TGCAGCAAGAGTAGGCGGATTCACAGCGGCGCCAGCCAGTACAGCCC 636
945 .....SerThrAlaAlaGlyAspAsnProValGlnThrArgTr 956
637 GAGCTGAGCAGATCGGCAATGCGCGGAGGATTTACAGCGACATGCGACA 686
956 oSerGlnProHisValThrMetProProLysSerSerGluProIleL 973
687 TATCGTCA.....AAACATCATCGCGCGGCGGAG 718
973 eIleuSerGluValLeuGlnProThrArgLeuProSerAlaThrThrSer 989
719 AAATTGCGCGCAGCGGATCGCTGACGGGTAAAGCGAAGGCTCAAC 768
990 GlnAlaIleProValThrGlnProIleArgHisProSerProProValAl 1006

```

```

769 ATGCTGTATGACAGCGCTTGCGTCTGCTTCCACCGCAAAAGAGATGC 818
1006 aThrValIleProThrAlaValAlaAspLysLysProValSerGlnsng 1023
819 GCGGATCAACGATT.....TGCGAGATATGGCGCAACAGTACATAG 862
1023 InGlySerAsnValProLeuPheAsnIleThrAsnSerSerAsnGlyTyr 1039
863 CCGGACGACGATCCGCGCATTTGGCAGTCCAAACCCCATATGCGGCA 912
1040 ProGln.....LeuAsnGlyTyrProAsnTyr 1048
913 GCGATAGAAGCGGTACGATATCTTTACGCGAGTCAATCCCGCTCAAG 962
1048 rGlyAsnGlyPheGlnAlaIleValGlyTyrGlyMetAsnTyrHisGln 1064
963 GATTGAGCTGTTCGGGA.....AAATACGCTTTGGCGCGCATCAGC 1006
1064 TyrProGlyTyrGlnGlyTyrAsnSerTyrGlyAsnGly..... 1077
1007 CACATCTGTGTAAGGCGTCCGAGATGGCGAGATGSCATTCGCGAAGG 1056
1078 .....MetGlyGlnLeuAlaIleThrHisAsn 1086
1057 AAATCGCGCTCAGCGACAAATTTCGCGATGCGCATCGCCAAATACC 1106
1087 .....AlaValThrSerLeuPr 1092
1107 GTCCCTTTCATCTCCGCAATATCCGTTCAACATGGAGCAGCTTACG 1156
1092 oProLeuValProSerGluAsnArgPheSerGlyThrAlaGlnProLeu 1109
1157 GCAAAGAAACATC..... 1170
1109 LysIleSerAspIleMetGluPheLeuGlyThrGlnGlnArgGlnAlaGly 1125
1171 ...ACCTCTCAACCTGCGCGCGTCAACGGAAG..... 1203
1126 SerSerSerArgAlaValProProAlaSerAlaSerThrSerAlaIaSe 1142
1204 .....AATGTGAACCTGGCAACAACAGCCAGCAAGACCAAG 1243
1142 rGlyIleThrAspLeuSerMetAlaAspLysMetGluValLeuTyrArg 1158
1244 TGCCGTTTGACGGTAAAGGTTTCCGATTTGAAAAGACGTAATAATC 1293
1159 .....GluAlaAspPheThrHis 1164
1294 GATACGAGAAATTAATACCGCTGACCAAGAGTGAAT 1329
1165 LysGlyAsnGlyAspThrMetValSerGlnCysAsn 1176

seq_name: p1r1:S48478
seq documentation block:
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N:Alternate names: extracellular glucosylase; mncin-like protein MDC1; protein YIR01
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C:Accession: S48478; A26877; B26877; S27281; J06123
R:Rowley, K.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48478
A:Accession: S48478
A:Molecule type: DNA
A:Residues: 1-1367 <R0K>
A:Cross-references: GB:Z47047; EMBL:Z38061; NID:9603997; PID:9763364; GSPDB:GN00009;
R:Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A:Title: Gene fusion is a possible mechanism underlying the evolution of STR1.
A:Reference number: A91831; M01D:87194600
A:Accession: A26877

```

A: Molecule type: DNA
 A: Residues: 1-242 <YAM>
 A: Cross-references: EMBL: M16164; NID: g172522; PIDN: AAA35014.1; PID: g172525
 A: Accession: B26877
 A: Molecule type: DNA
 A: Residues: 762-1331 <YAZ>
 A: Cross-references: EMBL: M16165; NID: g172523; PIDN: AAA35015.1; PID: g172526
 A: Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
 FEBS Lett. 239, 179-184, 1988
 A: Title: Similar short elements in the 5' regions of the STR2 and SGA genes from Sacchar
 A: Reference number: S27281; MIMD: 89031230
 A: Accession: S27281
 A: Molecule type: DNA
 A: Residues: 1-31 <PAR>
 A: Cross-references: EMBL: X13857; NID: g4551; PIDN: CAA32069.1; PID: g4552
 A: Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
 A: Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseudoty
 A: Reference number: J06123; MIMD: 96523237
 A: Accession: J06123
 A: Status: nucleic acid sequence not shown
 A: Molecule type: DNA
 A: Residues: 1-1367 <LAM>
 A: Cross-references: GB: U030626; NID: g1304386; PIDN: AAC49609.1; PID: g1304387
 A: Genetics: SGD
 A: Gene: SGD: MUC1; STR2; MAL5; DEX2; SGD: S0001458
 A: Cross-references: MIPS: Y1R019c; SGD: S0001458
 A: Map position: 9R
 C: Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
 F: 5-21/Domain: transmembrane #status predicted <TM1>
 F: 1350-1366/Domain: transmembrane #status predicted <TM2>

alignment_scores:
 Quality: 147.00 Length: 532
 Ratio: 0.546 Gaps: 18
 Percent Similarity: 50.564 Percent Identity: 20.865

alignment_block:
 US-09-303-518d-465 x S48478 ..

Align seg 1/1 to: S48478 from: 1 to: 1367

```

110 TCAGCGTCGATTCGACACCGGAGGAAATACCACTATTCGACGC 159
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
438 ThrThrGluSerSerSerAlaProValThrSerThrThrGluSer 454
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
160 AGGGGGAACTTGCAGCGGACGCGTCATATCGGATTTGGAAACATCA 209
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
454 rSerAlaProValProThrProSerSerSerThrThrGluSerSer 471
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
210 AACCCATCAGTTGGCAACCTGTTTCATCCAGAGCGGCGCATTAAGA 259
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
471 lalProValThrSerSerThrThrGluSerSerAlaProValProThr 487
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
260 ATATCGGTACATG.....TCGGTTTCGATACGCGGACGAA 300
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
488 ProSerSerSerThrThrGluSerSerSerAlaProValThrSerSer 504
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 GTTCATTCCTCCCTCGACACCATGCTCATTCGATTTCGATGAGC 350
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
504 rThrGluSerSerSerAlaProValPro.....ThrP 515
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
351 CGGTAGTCCCGTTGACGATTCAGCCTTAC.....GCA 385
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
515 rSerSerSerThrThrGluSerSerSerAlaProAlaProThrProSer 531
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
386 TCACATTCGACGATACGACACATCCGCGGCGGATGAGCGGCGCA 435
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
532 SerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGl 548
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
436 CAGGGCGGCGCT...ATCCGCTCCCAAGGCGGAGGATATATACAG 482

```

```

548 : .....||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    : userSerSerAlaProValProThrProSer.....SerS 560
483 CTCAGACATTAAGCGGTTGCCCAAAATATCCGCCATTCGACCGACA 532
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
560 erThrThrGluSerSerSerThrProValThrSerSerThrThrGluSer 576
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
533 ACCGACGA.....CCGACAAAGCGCTTTCGACCGCTTTCAC 570
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
577 SerSerAlaProValProThrProSerSerSerThrThrGluSerSer 593
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
571 ATACCGTAGTATGCTGACGACGAGAGTAGCGGACGATTAACGCGC 620
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
593 lalProValProThrProSerSerSerThrThrGluSerSerSerAlaP 610
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
621 CA...CCGATACAGCCCGAGCTGACGACATGCGGCAATGCGCGAAG 667
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
610 roAlaProThrProSerSerSerThrThrGluSerSerSerAlaProVal 626
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
668 CTTTCACGCGCACTGCATATGTC.....AAACATCATCGGC 708
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
627 ThrSerSerThrThrGluSerSerSerAlaProValProThrProSer 643
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
709 GCGGACAGAAATGTCGCGGACGCGGATGCGGACGAGATATACGCA 758
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
643 rSerThrThrGluSerSerSerSerAlaProValProThr.....ProS 657
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
759 AGGCTCAACATTCGTTATGACGCGGCTTGCGTCTGTTCCACCGAA 808
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
657 erSerSerThrThrGluSerSerSerAlaProValProThrProSerSer 673
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
809 ACAAGATGGCGGCGCATCAAG.....ATTGCGCATATGCGCACTC 852
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
674 SerThrThrGluSerSerSerSerAlaProValThrSerSerThrThrGl 690
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
853 AAAGATATGCGGACGACCATCCGATGCGGATGCGGACCAAAACCCCA 902
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
690 rSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProV 707
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
903 TGCCG.....CACAGGATAGAAAGCG 925
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
707 alProThrProSerSerSerThrThrGluSerSerSerAlaProValPro 723
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
926 TCAGCATATCTTTACGCGACATCCCGGCAAAAGATGAGCTGT 975
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
724 ThrProSerSerSerThrThrGluSerSerSerAlaProValProThrPr 740
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
976 CGGGAAATACGCGCTTGCGGATCAGCGACATCTGTCAGACGGTC 1025
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
740 oSerSerSerThrThrGluSerSerSerSerAlaProValThrSer..... 755
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1026 GCAGATGGCGCAGATCGATTGCGGAAAGGAAATCCGCGCTGACGACA 1075
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
756 .....ThrThrGluSerSerSerAla 762
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1076 ATTTGCGGATGCGGACATACGCAAAATACCGTCCCTTACCATTCGCA 1125
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
763 ProValPro.....ThrProSerSerSerThrThrGluSerSer 776
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1126 AATTCGTTAACTTGAGACGCTTACGCGCAAAAGAAACATCACCTC 1175
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
776 lalProValProThrProSerSerSerThrThrGluSerSerSerAlaP 793
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1176 CTCACCGTG..CCGCGTCAAAAGGAAAGAAAGTGAAGATGCGCAACAA 1224
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
793 roValProThrProSerSerSerThrThrGluSerSerSerAlaProVal 809
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1225 CGCCACCGAAGACCAAGTCCGTTTGACGTTAAAGGTTTCCGAATT 1274
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
810 ProThrProSerSer..... 814
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1275 TGAAAAAGAGGTAAATACGATACGAAATTAATACCGCTGATCCGCAAG 1324
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

815 .....SerSerAsnIleThrSerSerAlaProSerS 825
1325 TGAATCCTATAGATGAA.....CCCGCTCTTAAAT 1353
      |||||:|||||:
825 erThrProPheSerSerSerThrGluSerSerSerValProValProThr 841
1354 CCTAAGAGTTCTGTC.....GGATCGGCTCATCTTGCTCTAATAC 1394
      |||||:|||||:
842 ProSerSerSerThrThrGluSerSerSerAlaProValSerSerThr 858
1395 TGGCAGATTCATATACGAAATATACCAAGGACAGTAGAATCAGATATA 1444
      |||||:|||||:
858 rThrGluSerSerValAlaProValProThrProSer..... 870
1445 TCCGACCACTAAATTAATCTCTCTTACGACCGCTACCAAAAGACCTAAT 1494
      |||||:|||||:
871 ..SerSerSerAsnIleThrSerSerAlaPro...SerSerIleProPhe 885
1495 AATGATATTTGGATTAATTTGGTAATGATGACTTAAGGTCATCAAG 1544
      |||||:|||||:
886 SerSerThrThrGluSerPheSerThrGlyThrThrValThrProSerSe 902
1545 AACTAAGTCACAGATTGGAATGGGATGTTCAATGTCTAATAACA 1590
      |||||:|||||:
902 rSerIleTyrProGlySerGlnThrGluThrSerValSerSerThr 917
seq_name: p1r2:T29634

```

```

seq_documentation_block:
hypotheetical protein C12D12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T29634
R:Nhan, M.; Hawkins, J.
submitted to the EMBL Data Library, March 1996
A:Description: The sequence of C. elegans cosmid C12D12.
A:Reference number: Z20556
A:Accession: T29634
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1 825 <NHA>
A:Cross-references: EMBL:U51998; PIDN:AA96080.1; GSPDB:GN00028; CESP:C12D12.1
A:Experimental source: strain Bristol N2; clone C12D12
C:Genetics:
A:Gene: CESP:C12D12.1
A:Map position: X
A:Introns: 48/1; 86/3; 137/1; 172/3; 224/3; 253/1; 287/3; 328/2; 454/1; 487/3; 692/1
C:Superfamily: Epstein-Barr virus membrane antigen gp350

```

```

alignment_scores:
Quality: 146.50      Length: 401
Ratio: 0.763        Gaps: 18
Percent Similarity: 47.880      Percent Identity: 22.195

```

```

alignment_block:
US-09-303-518D-465 x T29634 ..

```

```

Align seg 1/1 to: T29634 from: 1 to: 825

```

```

281 TTTCGATCAG.....GGCAGAGTCATCCCTTCGACAACCAT 324
      |||||:|||||:
470 TTPProValThrThrValSerThrMetSerProThrThrThrValThrVa 486
325 GCCTCAGATTCGATTCGATG.....AAGCGGTAG 356
      |||||:|||||:
486 lProThrThrProThrProValProThrThrThrThrThrProProAla 503
357 TCCCGTTCGAGATTCAGCCTTACCGATTCATGGAGGATACGAC 406
      |||||:|||||:
503 snProThrThrAlaThrProThrThrValGlyThrSerIleGlnThrAsn 519
407 ACCATCCCGCG.....ACGGCTATGACGGGACACAGGGCGGCTAT 450

```

```

|||||:|||||:
520 ThrIleSerProHisLeuSerThrThrIleThrGlySerIleValThrSerTh 536
451 CCCGCTCCCAAGGCGCCAGGATATATACAGCTTCGACATATAAAGGCGT 500
      |||||:|||||:
536 rProThrMetAlaProGlnThrThrSerAlaSerProThrThr.....T 550
501 TGGCCCAAAATATCCGCTCAACCTGACCGACACACCGACACCGGACAC 550
      |||||:|||||:
550 hProThrThrIleThrThrAlaSerGlnProThrThrThrLysProValVal 566
551 GCGTGTGACCGTT.....TCCCAATATACCGGTAGT 582
567 ThrThrAsnSerValThrProSerThrGlyThrThrThrThrValProValPr 583
583 ATGTGACGCCAAGAGTAGCGACGATTCACACCGCA.....CCCG 626
583 oThrThrThrGlySerProThrThrGlnThrThrAlaProValThrLysP 600
627 ATACAGCCCCGAGCTGGACAGATCGGGCAATCGCCGACGATTTCACAG 676
      |||||:|||||:
600 roThrValProSerSerThrThrGlnThrAlaProPro...ValThrThr 615
677 GCACCTGACATATCGTCAAAACATCATCGGCGGCGGACGAGAAATGTC 726
      |||||:|||||:
616 ProThrSerGlnProProValThrThrThrThrSerLeuThrThrLeuTh 632
727 GCGCAGCGGATGCGCTGACGAGGTATAGCGAAGGCTCAACATGCTGT 776
      |||||:|||||:
632 rThrProThrValProVal.....ThrThrThrValValP 644
777 TATGACAGCGCTGGGTC...TGCTTCCACCGAAACAGATGGCGCGCA 823
      |||||:|||||:
644 roSerSerAlaThrValProThrThrProProThrThrThrValThrValAla 660
824 TCAACGATTTGGCAGATATGGCGCACTCAAGACTATGCGCGACGCC 873
      |||||:|||||:
661 AlaThrThrThrSer.....LysAlaProValValThrThrSerPr 674
874 ATCCGCGATTGGCGAGTCCCAAAACCCCAATGCCGACACAGCATTAAGC 923
      |||||:|||||:
924 CGTAGCAATATCTTTACGCGACTCATCCCGTCAAAAGGATGGAGCTG 973
      |||||:|||||:
689 roSerThrValGly.....ThrSerProThrAlaProAlaAsnLeu 702
974 TTCGGGGAATAATACGCGCTGGCGGACATCAGCACATCTGTCAAGCG 1023
      |||||:|||||:
703 ThrThrProThrThrAla..... 708
1024 TCGCAGATGGCGGAGATCGCATTTGCCGAAGGGAATCCGCGTCAGCA 1073
      |||||:|||||:
709 .....ProValAsnProThrSerSerT 716
1074 CAATTTGCGATGCGG.....CATAGGCCAAATACCGCTGCC 1111
      |||||:|||||:
716 hThrAlaProThrAlaProValAsnProThrSerProThrThrAlaPro 732
1112 CTATACATTCGCGAATATCGCTTCAAACTTGACAGCAGCTTACGGGAAA 1161
      |||||:|||||:
733 ...ThrValProProValThrThrThrThrProThrThrThrThrSerTh 748
1162 GAAACATCAGCTCTCAACGCTGCGCGCGCGCAACGAAAGAAATGGA 1211
      |||||:|||||:
748 rThrThrThr...ThrThrThrThrThrThrThrThrThr..... 758
1212 ACTGGAACAACAAGCCACCGCAAGAACCAAGTGCCTTTGACGGTAAG 1261
      |||||:|||||:
759 .....ThrGlnThrThrThrProThrThrProValVal..... 768
1262 GGTTCGAAATTTGAAAGAGAGTAATAATACGATACGAGATTAATAC 1311
      |||||:|||||:

```



```

454  InProValpProGlyThrLysArgThrLysSerAsnTyrcysGlu..... 468
1153  TACGGCAAGAAAACATCACCTCTCAACCGTCGCCCGCTCAACAGGAAA 1202
469  ....AsnGluProAsnLysSerSerGlnSerLeuValSerSerGluSerH 484
1203  GAATGTGAACCTGGCAACCAACGCCACCGAGACCAACAAAGTCGGCTTG 1252
484  sAsnValGluGlyThrAsnTyrcysGluThrGlyThr..... 496
1253  ACGGTAAAGGTTTCCGAATTTGAAAAAGACGTAATAATACGAT..... 1296
497  ....ValGlyPhe.....TyrAspProSer 503
1297  ACGAATTTATACCGCTGTACCAACAGTGAATCATATGATCAACCCCT 1346
504  AlalulSerAlaSerIleAspGluLeuArgGln...SerThrProVal 519
1347  CTTTAATCTTAAGGTTCTGCGATCGGCTCATCTTGTGCTATTAATCG 1396
519  lAlaArgAspSerGluLeuLeuSerArgAlaHisSerPheAspLeuAsn. 535
1397  CCAGAATTCAATACGCAAAATTTACCAAGCAGAGTAGAATCAGATATATC 1446
536  ..ArgLeuAsp.....Leu 539
1447  CCACCTTAATAATTAATCTCTTCAGCAGCGCTACCAAAAGACATAATTA 1496
540  ProSerGlnAspLysSerThrSerTyrgluValProAsnGlyThrGluAs 556
1497  T.....CGATATTTGGATAAA...TTTG 1516
556  ngInSerProArgProValThrSerLeuGlyPheValAsnGluThrPheP 573
1517  GTAATGATGACTAATAAGCTCCA 1539
573  heGluGluLysProLysAlaPro 580

seq_name: p1r2:JT0345

seq_documentation_block:
  description (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)
  N:Alternate names: sucrose 6-glucosyltransferase
  C:Species: Streptococcus mutans
  C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Sep-1999
  C:Accession: JT0345, C33135
  R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.
  Gene 69, 101-109, 1988
  A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.
  A:Reference number: JT0345; MUID:89137980
  A:Accession: JT0345
  A:Molecule type: DNA
  A:Residues: 1-1375 <UED>
  R:Experimental source: GS-5
  R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
  J. Bacteriol. 169, 4263-4270, 1987
  A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
  A:Reference number: A33135; MUID:87308013
  A:Accession: C33135
  A>Status: Preliminary
  A:Molecule type: DNA
  A:Residues: 1-349 <SHI>
  A:Cross-references: GB:M17361
  C:Genetics:
  A:Gene: gtfC
  C:Function:
  A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans
  C:Superfamily: cpl repeat homology
  F:Keywords: duplication; glycosyltransferase; hexosyltransferase
  F:1-34/Domain: signal sequence #status predicted <SIG>
  F:35-1375/Product: glucosyltransferase #status predicted <ANT>
  F:1126-1145/Domain: cpl repeat homology <CP1>

```

```

F:1253-1272/Domain: cpl repeat homology <CP2>
F:1318-1337/Domain: cpl repeat homology <CP3>

alignment_scores:
  Quality: 140.00      Length: 606
  Ratio: 0.507        Gaps: 32
  Percent Similarity: 45.545      Percent Identity: 22.112

alignment_block:
  us-09-303-518d-465 x JT0345

Align seg 1/1 to: JT0345 from: 1 to: 1375

64  GCACAGCCCTCAGATTGTGGCAACGATTTGTTATCGGACAGTTCTGCA 113
    :::::|||||  ||:::|||||  ||  :::::|||||
643  SerTyrlAlaLeuLeuLeuThrAsnLysSerSerValProArgValTyrr 659
    :::::|||||  ||:::|||||  ||  :::::|||||
114  CCGTCACGATTTGCAACCCGACGGGAATAC..... 144
    :::::|||||  ||:::|||||  ||  :::::|||||
659  rGlyAspMetPheThrAspAspGlyIntTyrrMetAlaHisLysThrIleA 676
    :::::|||||  ||:::|||||  ||  :::::|||||
145  .....CACCTATTC 153
676  snTyrgluAlaIleGluThrLeuLeuLysAlaArgIleLysTyrrValSer 692
    :::::|||||  ||:::|||||  ||  :::::|||||
154  GGCACAGGGGGGGAATCTCCGACGCGCGTCAT..... 189
    :::::|||||  ||:::|||||  ||  :::::|||||
693  GlyGlyGlnAlaMetArgAsnGlnGlnValGlyAsnSerGluIleLeTh 709
    :::::|||||  ||:::|||||  ||  :::::|||||
190  .....ATCGATTGGGAACATACAAAGCCATCAGTTGGGCAAC 229
    |||  |||  :::::|||||  ||  :::::|||||
709  rSerValArgTyrglyLysGlyAlaLeuLysAlaThrAspThrGlyAspA 726
    :::::|||||  ||:::|||||  ||  :::::|||||
726  rGthrThrArgThrSerGlyValAlaValIleGluGlyAsnAspProSer 742
    :::::|||||  ||:::|||||  ||  :::::|||||
271  ATP.....GTCCGCTTTCCGAT.....CA 290
    :::::|||||  ||:::|||||  ||  :::::|||||
743  LeuArgLeuLysAlaSerAspArgValAlaValAsnMetGlyAlaAlaH 759
    :::::|||||  ||:::|||||  ||  :::::|||||
291  CGGGCAGCAAGTCATTCGCCCTTC.....GACACAC.....C 322
    :::::|||||  ||:::|||||  ||  :::::|||||
759  slyAsnGlnAlaTyrrArgProLeuLeuLeuThrThrAspAsnGlyLe 776
    :::::|||||  ||:::|||||  ||  :::::|||||
323  ATGCCACATTCGATCTGTGATGAAGCCGCTAGTCCGTTGACGAGATTC 372
    |||  |||  ||:::|||||  ||  :::::|||||
776  ySalArgTyrrHisSerAspGlnGluAlaIleGly..... 786
    :::::|||||  ||:::|||||  ||  :::::|||||
373  AGCCTTTACCGCATTCATTGGAGCGATACGAA.....CACCATCCGC 416
    |||  |||  |||  |||  |||  |||
787  ..LeuValArgTyrrThrAsnAspArgGlyGluLeuIlePheThrAlaAl 802
    :::::|||||  ||:::|||||  ||  :::::|||||
417  CGAC.....GGCTATGACGGGCGACAGGGGGGCTATCC..... 453
    |||  |||  ||:::|||||  ||  :::::|||||
802  aAspIleLysGlyTyrrAlaAsnProGlnValSerGlyTyrrLeuGlyVal 819
    :::::|||||  ||:::|||||  ||  :::::|||||
454  ..GCTCCCAAGGGCGGAGGATATATACAGCTACGACATTAAGCGCTT 501
    :::::|||||  ||:::|||||  ||  :::::|||||
819  rValAlProValGlyAla.....AlaAla 826
    :::::|||||  ||:::|||||  ||  :::::|||||
502  GCCCAAAATATCCGCTCAACCTGACCGACAGACCGACGCGAGCAAG 551
    |||  |||  |||  |||  |||  |||
827  AspGlnAspValArgValAlaAlaSerThrAlaProSerThrAspGly 843
    :::::|||||  ||:::|||||  ||  :::::|||||
552  GCTTGTCACCGTTTCCACATACCGGTAGTATGCTGACCAAGAGATGAG 601
    :::::|||||  ||:::|||||  ||  :::::|||||
843  sSerVal.....HisGlnAsnAlaAlaLeuAspSerArgValMetP 857
    :::::|||||  ||:::|||||  ||  :::::|||||
602  GCGACGATTCMAAGCGCACCGCATACAGCCCGGCGGTGACAGATGCG 651
    :::::|||||  ||:::|||||  ||  :::::|||||
857  heGluGlyPheSerAsnPheGlnAlaPheAlaThrLysLysGluGluTy 873
    :::::|||||  ||:::|||||  ||  :::::|||||

```

```

652 GGCAATGCCCGAGAGCTTTCAC.....GGCAC 680
      |||:|||||
874 ThrAsnValValIleAlaLysAsnValAspLysPheAlaGluProIleValThr 890
      |||:|||||
681 TGCAGATATGTCATAAAACATCATCGCGCGGAGAGAAATGTGTGGCG 730
      |||:|||||
890 LThrAsp.....PheGluMetAlaProGluIntYrValSer 902
      |||:|||||
731 CAGGCGATGCGGTG.....CAGGCTATTAAGCAAGAGCTCAAACTAT 771
      |||:|||||
902 eThrThrAspGlySerPheLeuAspSerValIleGluAsnGlyTyrAlaPhe 918
      |||:|||||
772 GCTGTATGCAAGCGCTTGGCTGTGCTTTCACCGAAACAAAGATGGCGC 821
      |||:|||||
919 ThrAspArgTyrAsnProIleGly.....IleSerLysProAsnLysTyrGly 934
      |||:|||||
822 CATCAACGATTTGGCAGAT...ATGGCGCAACTCAAGACTATGCGCGCAG 868
      |||:|||||
934 rAlaAspAsnLeuValLysAlaIleLysAlaLeuHisSerLysGlyIleL 951
      |||:|||||
869 CAGCATCCCGGATTTGGGAGTCACAAACCCCAATGCCGACACAGCATATA 918
      |||:|||||
951 ySvalMetAlaAspTyrValProAspGluMetTyrAlaLeuProGluLys 967
      |||:|||||
919 GAAGCGCTCAGCAATATCTTTACGGCAGTCACTCCCGTCAAGAGGATTCG 968
      |||:|||||
968 GluValValThr..... 971
      |||:|||||
969 AGCTGTTCG...GGAAATACGGCTTGGGCGCATCAGCAGCATCTGT 1015
      |||:|||||
972 .AlaThrTyrValAspLysTyrGlyThr.....ProVal 982
      |||:|||||
1016 TCAGCGGTGCGCAGATG...GGCAGATCGCATTTGCCGAAGGAATCC 1062
      |||:|||||
982 aLAlaGlySerGluIleLysAsnThrLeuTyrValValAspGlyLysSer 998
      |||:|||||
1063 GCCGTCACCGCAAT.....TTTCCGATTCGCGC 1091
      |||:|||||
999 SerGlyLysAspGluGlnAlaLysTyrGlyValAlaPheLeuGluGluLe 1015
      |||:|||||
1015 uGlnAlaLysTyrProGluLeuPheAlaArgLysGluLe..... 1028
      |||:|||||
1142 TGGACAGCGCTTACGGCAAAACATCACTCTCTCAACGTCGCGCG 1191
      |||:|||||
1029 .....SerThrGlyValProMet 1034
      |||:|||||
1192 TCAGACGGAAGATGTGAA...CTGGCAACAAACGCCACCGCAAGAC 1238
      |||:|||||
1035 AspProSerValLysIleLysGluIleTyrPheAsnGlyThr 1051
      |||:|||||
1239 CAAAGTCCGTTGACGGTAAAGGTTTCCGATTTGAAAAAGACGTAA 1288
      |||:|||||
1051 rAsnIleLeuGlyArgGlyAlaGlyTyr.....ValL 1062
      |||:|||||
1289 AATAGATACGAGAAATTATACCGCTGTACCAAGATGATCTATTAAT 1338
      |||:|||||
1062 euLysAspAlaIleThrAsnThr.....TyrPheSerLeuValSer 1075
      |||:|||||
1339 GAACCGCTTTAATCTTAAGGTTCTGTGCGATGGCTCAT.....TC 1382
      |||:|||||
1076 AspAsnThrPheLeuProLysSerLeuValAsnProAsnHisGlyThrSe 1092
      |||:|||||
1383 TTGCTTAAATCTCCAGAAATTCATACCAAAATTAACAGCAAGATA 1432
      |||:|||||
1092 rSerSerValThrGlyLeuValPheAspLys.....GlyT 1105
      |||:|||||
1433 GAATCAGATATATCCACCTAAATTAATCTCTTCACGACCGCTACCA 1482
      |||:|||||
1105 yrValIlyrIySerThrSerGlyAsnGlnAlaLysAsnAlaPheIleSer 1121

```

```

1483 AAAGCACTTAATATGATATTGCAATAATTGTATGATGACTAA 1532
      |||:|||||
1122 LeuGlyAsnAsnTyrTyrPheAspAsnAsnGlyTyrMetValThr.. 1137
      |||:|||||
1533 AGGTCCATCAAGAACTAAAGCTCAAGAAATTTGAATGC.....GATG 1573
      |||:|||||
1138 .GlyAlaGlnSerIleAsnGlyAlaAsnTyrTyrPheLeuSerAsnGly 1154
      |||:|||||
1574 TTCATTTGTCTAAA.....ACAGAGACGACACTTGATGAG 1611
      |||:|||||
1154 LeuIleuArgAsnAlaIleTyrAspAsnGlyAsnLysValLeuSerTyr 1170
      |||:|||||
1612 GCTAGTAGGATGTATAG 1629
      |||:|||||
1171 TyrGlyAsnAspGlyArg 1176
      |||:|||||
seq_name: p1r2:J00405
seq_documentation_block:
hypothetical 119.5K protein (uvra region) - Micrococcus luteus
N:Alternate names: ORF 1 protein
C:Species: Micrococcus luteus, Micrococcus lysodeikticus
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 04-Feb-2000
C:Accession: J00405
R:Shiota, S.; Nakayama, H.
Mol. Gen. Genet. 217, 332-340, 1989
A:Title: Micrococcus luteus homolog of the Escherichia coli uvra gene: identification
A:Reference number: S04781; MUID:89364717
A:Accession: J00405
A:Molecule type: DNA
A:Residues: 1-1106 <SH1>
A:Cross-references: EMBL:X15867
A>Note: this reading frame extends between two stop codons and does not begin with a
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
alignment_scores:
Quality: 139.50 Length: 638
Ratio: 0.524 Gaps: 37
Percent Similarity: 41.693 Percent Identity: 24.295
alignment_block:
US-09-303-518D-465 x J00405 ..
Align seg 1/1 to: J00405 from: 1 to: 1106
30 TCTGTCCATCTGCGAGTGTGCTGCCGATGCATGCACACGCCCTCAGATT 79
      |||:|||||
116 SerValArgLeuGlnValLeuProGlyLeuArgHisAspLeuArgAr 132
      |||:|||||
80 TGGCAACGATTTCTTTATCCGCGCAGGTCTCGACCGTCAACGATTCGAA 129
      |||:|||||
132 gGly.....ProAlaIleuArgValAlaLeuLeuL 144
      |||:|||||
130 CCGGACGGAATATACACCTATTTCGCGACGAGGGGAGAACTTCCGAGCG 179
      |||:|||||
144 euArgProHisValProGlyProGlyGlyGlnAlaGlyArg...GlyLeu 159
      |||:|||||
180 CAGCGGTATATCGGATTTGGGAAACATACAAAGCATCACTGTGGCAACC 229
      |||:|||||
160 HisArgGlyProValProGlyArgValHisArgProIleValHisGlnPr 176
      |||:|||||
230 TGTTCATCCGACA.....GGCGGCATTAAGGAATATATCGGCTAC 270
      |||:|||||
176 o...GlnProAlaLeuHisSerGlyAspHisHisArgAspLeuArgLeuH 192
      |||:|||||
271 ATGTCCGCTTTTCGATCA..... 290
      |||:|||||
192 iSAlaProAlaLeuGlyThrCysArgGlyAlaAlaLeuProAlaValArg 208
      |||:|||||
291 CGGCGACGAACTCATT.....CCCTTCGACA 319
      |||:|||||

```

```

209 ArgAlaGlyGluProAlaAspProAlaAlaAspArgGlyProAlaArgGly 225
320 ACCATGCTCATCATTCGATCTGTGATGAAGCGGTAGTCCGTTGACGGA 369
    ::::
225 yAlaAla.....ArgA 229
370 TTCAGCGCTTA.....CCGATCCATTGGGACGG 398
    ::::
229 LAlHsProLeuProGlyAlaArgGlyProArgProGlnGlyArg 245
399 ATACGACACCATCCCGCGAGCGGTATGACGGCCACAGGCGCGCGCT 448
    ::::
246 ValArgGlyProValGlnGlyProVal...HisAlaGlyLeuArgArgArg 261
449 ATCCGCGCTCCAAAGG.....CGCGAGG 471
    ::::
261 GgLyArgGlyAspArgProAlaLeuGlyProAlaArgAlaGlnGlyAlaG 278
472 GATATATACAGCTACGACATAAAGCGCTTGCCAAA..... 509
    ::::
278 LyGlnAlaHisHisArgArgArgGlyProProArgHisGlnGlyGly 294
510 TATCGCGCTCAACCTGACCGACACCGGACGCGGACAGCGCTTGCG 559
    ::::
295 HisProAlaPro.....HisArgLeuGlyGlyAspArg 306
560 ACCGTTTCCACATACCGGTAGTAGTCGACGACGAGTAGCGACGGA 609
    ::::
306 gPro.....GlnAlaGlyGlyArg..... 312
610 TTCMAACGCCACCCGATACGCCCGGAGCTGACAGATC..... 650
    ::::
313 .....ProArgArgGlyArgValArgGlyArg 321
651 GGGCAATGGCGCGGAGCTTTCACAGCGACCTGAGATATTCGCAAAACA 700
    ::::
322 GLyAlaGlyArgArgGlnGlyGlnGlnHisArgGlyValArgArgAl 338
701 TCATGCGCGCGGCG.....AGGAGAAATTTGTC 726
    ::::
338 aGlyArgArgGlyGlnProProValProLeuValLeuArgGlnAlaLeuL 355
727 GCGCGAGCGGATCGCGTGCAGGGTATTAAGCCGAGCTCAACATTGCTGT 776
    ::::
355 euProGlnArgAlaArgAla..... 361
777 TATGACGCGCTGGGCTTGCTTCCACCGAAACAAGATGGCGG..... 821
    ::::
362 .....AspArgGlyArgAspArgAlaAlaLe 370
822 .....CATCAACGATTTGGCAGATATGGCGCAACTCAAGACTAT 861
    ::::
370 uValLeuLeuGlnGlnProValArgArg.....ValProArgValH 384
862 GCGCGACGACGATCCGCGATTTGGCAGTCCAAACCCCAATGCCGACA 911
    ::::
384 LsArgHisArgLeuProProAlaGlyGlyProGlyProArgArgArgGln 400
912 AGGCAT.....AGAACCGCTGACGAAATCTTTACGGGACGCA 949
    ::::
401 ArgArgAlaValAlaProAlaArgGlyArgArgArgAlaValAlaArgGly 417
950 TCCCCGTCAAAGGAGATGGAGCTGTTCCGGGAAAA.....TACGGC 990
    ::::
417 nValHis...LeuGlyLeuLeuAlaAlaArgAlaArgAlaGlyGlnGly 433
991 TTGGCGGCATCAGCGGACATCTCTCAAGCGGT..... 1024
    ::::
434 AspGlyLeuLeuProGlnHisProValGln...GlyProAspGlyGlyGly 450
1025 .....CGCAGATGGCGAGATCGCATTCGCGAAAGCGAAAT..... 1060
    ::::
450 LArgArgArgProAlaArgGlnGlyLeuGlnGlyGlyAspValPro 466

```

```

1061 CCGCGCTCAGGACATTTTGGCGATGGGATACCGCAATACCGGT... 1108
    ::::
467 GlnProValArgProAlaArgAlaProLeuHisHisGlyLeuArgGlyArgH 483
1109 .CCCGTTACCATTTCCGAAATATCC..... 1132
    ::::
483 sProLeuArgAspArgAlaGlnAlaArgGlyAspArgValGlyArgArgPro 500
1133 .....GTTCAAACTTGGACGACGCTTACCGCAAGAAACA 1168
    ::::
500 rGAlaLeuArgValAlaHisAlaGlyAspProValProGlyValProArg 516
1169 TCACCTCTCCACCGCTG.CCGCGCTCAAGCGAAAGAAATGTAACGCGC 1217
    ::::
517 GLyProProGlnProHisGlyProGlnArgAlaArgGlyArgProValH 533
1218 AATCAACGCCGCCCGCAAGACAAA..... 1242
    ::::
533 sArgGlyArgHisProProAlaHisAlaArgGlyHisGlyValLeuLeuG 550
1243 .GTGCGCTTTGACGCTTAAGGCTT.....CCGAATTTTGA 1278
    ::::
550 LyAlaAlaAlaAspGlyProGlyAlaAlaAspArgGlyProGlyAlaGln 566
1279 AAAGAC.....GTAAATACGATACGAGATTAATACCGC 1313
    ::::
567 GLyAspProGlyProAlaGlyValProAlaGlyArgArg.....ProArg 581
1314 TGATACCAAGTAGTAATCTATAGATGAACCGCTTTAATCTCAAGGTT 1363
    ::::
581 gValProGlnProArgGAlaAlaGlyArg.....HisProLeuArgA 595
1364 CTGTCGATGGCGCTCATTTGGCTATATCTGCGCAATTCATCAATGCA 1413
    ::::
595 rGArgGlyProAlaHisProProGlyHisHisArgArgLeuArgAlaGly 611
1414 AAATTCACCAAGCGCAAGGTAGATCAGATATATC..... 1446
    ::::
612 ArgArgProLeuArgProArg...ArgAlaValHisArgProAlaProAl 627
1447 .....CCACCTAAATAATCTCTCTTCAGCACCGGCTCAACAAAGAC 1489
    ::::
627 aGlyGlnProProHisArgAspPro.....ProAlaProAlaGly 642
1490 CT.....AATTAAT 1497
    ::::
642 rOArgGlnHisProHisArgArgAlaArgAlaArgGlyHisAspArgArg 658
1498 GGATATTTGGATAAATTTGTAATGAATGAGCTAAAGCTCATCAAGAC 1547
    ::::
659 GLyGlyLeuAspArg...GlyHisArgProSerArgGlyArgValArgArg 674
1548 TAAAGTCAA 1557
    ::::
674 gArgGlyArg 677
seq_name: plr2:A48018
seq_documentation_block:
mucin 7 precursor, salivary - human
N:Alternate names: mucin, MG2; mucin, MG2a-T1; mucin, MG2b-T2
C:Species: Homo sapiens (man)
C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 07-May-1999
C:Accession: A48018; S29115; S29116; S29114
R:Bobek, L.A.; Tsai, H.; Biesbrock, A.R.; Levine, M.J.
J. Biol. Chem. 268, 20563-20569, 1993
A:Title: Molecular cloning, sequence, and specificity of expression of the gene encod
A:Reference number: A48018; MUID:9338636
A:Accession: A48018
A:Molecule type: mRNA
A:Residues: 1-377 <BOB>
A:Cross-references: GB:L13283

```

A:Experimental source: submandibular gland
 A:Note: sequence extracted from NCBI backbone (NCBI:137719, NCBI:137720)
 R:Reddy, M.S.; Bobek, L.A.; Haraszthy, G.G.; Biesbrock, A.R.; Levine, M.J.
 Biochem. J. 287, 639-643, 1992
 A:Title: Structural features of the low-molecular-mass human salivary mucin.
 A:Reference number: S29114; MUID:93075006
 A:Accession: S29115
 A:Molecule type: mRNA
 A:Residues: 143-168 <RED>
 A:Accession: S29116
 A:Molecule type: protein
 A:Residues: 'S', '71-79', 'N', '81-86', 'XX', '89', 'X', '91', 'P' <RE2>
 A:Accession: S29114
 A:Molecule type: protein
 A:Residues: 143-145, 'X', '147', 'XXX', '151-152', 'X', '154-158', 'X', '160-161', 'A', '163-164', 'XX', '167-168'
 C:Genetics:
 A:Gene: GDB:MUC7
 A:Cross-references: GDB:138799; OMIM:158375
 A:Map position: 4q13-4q21
 C:Keywords: glycoprotein
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-377/Product: mucin 7, salivary #status predicted <MAT>
 F:97,128,135,146,312/Binding site: carbohydrate (asn) (covaleqt) #status predicted

alignment_scores:
 Quality: 139.00 Length: 370
 Ratio: 0.952 Gaps: 19
 Percent Similarity: 39.459 Percent Identity: 24.324

alignment_block:
 US-09-303-518D-465 x A48018 ..

Align seg 1/1 to: A48018 from: 1 to: 377

```

119 AGCATTTCGACCCGAGGGGAAATACCACTATTGCGACGAGGGGGA 168
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
99 SerValValAsnProThrLeuValAlaThrGlnIleProSerValTh 115
169 CTTCGGCGGAGGGGTCATATGCGATGGGAACATCAAGCAATCA 218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
115 rPheProSerAlaSer.....ThrLysIleThrLeuProAsnValT 130
219 GTTGG.....GCACCTGTCATCCAGCAGGGGCGCATTAAG 256
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
130 hPheLeuProGlnAsnAlaThrThrIleSerSerArgLysAsnValAsn 146
257 GAAATTCGGCTACATGTCGGCTTCGATCAGGCGACCAAGTCCAT 306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
147 ThrSerSerValAlaThrLeuAlaProVal.....AsnSerProAl 161
307 TCCCCCTTCGACAAAC...ATGCTTCACATTCGATTCGATGAAGCGG 353
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
161 aProGlnAspThrThrAlaAlaProProThrProSerAlaThrThrProA 178
354 TAGTCCGTTGACGAGTACGCTTACCGCATTCATTCGAGGATACG 403
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
178 laProPro.....SerSerSerAlaProProGluThr 188
404 AACACATCCCGCGCGGCTATGACGGGCCACAGGGCGGCGGTATCC 453
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
189 ThrAlaAlaProProThrPro..... 195
454 GCTCCCAAGGGCGGAGGATATATACAGCTACGACATAAAGCGTTGC 503
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
196 .....SerAlaThrThr.....GlnAlaP 202
504 CCAAAATATCCGCTTCACCTACGACGACCGACCGACGACGAGC 553
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
202 rProSerSerSerAlaProProGluThrThrAlaAlaProProThrPro 218
554 TTTCGACCGTTCCGACATACCGGTATGCTGACGCAAGAGATAGCG 603
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

219 ProAlaThrThrProAlaProPro..... 226
604 GACGATTCGAAGCGCGCCACCGCATACA.....GCCCGAGCTGGACAG 647
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
227 .....SerSerSerAlaProProGluThrThrAlaAlaProProThrPro 242
648 ATCGGCAATGCGCGCGACGCTTCACAGGCGACTGCAGATATCGTCAAA 697
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
242 eAlaThrThrProAlaProLeuSer..... 250
698 ACATCATCGGCGCGCGAGGAAATGTCGCGCGACGCGATGCGTCGAG 747
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 .....SerSerAlaProProGluThrThrAlaAlaProProThrPro..... 264
748 GGTATTAAGCGAGCTTCACATTCGCTATGACAGCGCTTGCTGCT 797
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
265 .....SerAlaThrThrLeuAspProSerSerAla.....Sera 276
798 TTCACCGCAAAACAMGATGCGCGCATCAACGATTTGGCAGATATGCGC 847
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 laProProGluThr..... 280
848 AACTCAAGACTATGCGCGACGACCATCCGCGATTGGGATGCAAAAC 897
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
281 .....ThrAlaAlaProProThrProSerAla.....ThrTh 291
898 CCCAATGCGCGACAAAGCATAGAAAGCGTCAGCAATATCTTTACGGAGT 947
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
291 rProAlaPro.....ProS 296
948 CATCCCGCTCAAAAGGATTTGAGAGCTGTTCGGGAAATACGCGTGGCG 997
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
296 eSerProAla..... 299
998 GCATCAGGCGACATCTGTCAACGCGTCGCAAGATGGCGGATCGCAT 1047
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
299 ..... 299
1048 CCGAAAGGAATCCGCCGTCAGGACGACATTTCCGATGCGGACATACG 1097
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 .....ProGlnIleThrThrAlaAlaProIleThr.....ThrP 311
1098 CAATACCCGCTCCCTTACCATTT...CCGCAATATCCGTT..... 1135
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
311 roAsnSerSerProThrThrLeuAlaProAspThrSerGluThrSerAla 327
1136 .....CAACTGGAGCAGCGCTTACGCAAGAAAGAAACATCACC 1173
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
328 AlaProThrHisGlnThrThrThrSerValThrThrGlnThrThrThr 344
1174 TCCTCAACCG 1183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
344 rLysGlnPro 347
seq_name: p1r2:T02345
seq_documentation_block:
hypothetical protein KIA0324 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C:Accession: T02345
R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D
re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Pawcett, J.; Deaven, L.
submitted to the EMBL Data Library, March 1998
A:Description: Sequencing of human chromosome 16p13.3.
A:Reference number: Z14664
A:Accession: T02345
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1791 <RIC>
A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AAC08453.1; PID:g2996650
C:Genetics:
A:Map position: 16

```


alignment_scores: Quality: 137.50 Length: 572

 Ratio: 0.476 Gaps: 32

 Percent Similarity: 50.524 Percent Identity: 21.503

alignment_block:
US-09-303-518D-465 x T39758

Align seg 1/1 to: T39758 from: 1 to: 615

```

43 GCACGTGCTCCGCGATGATGCACACGCGCTGACATTTGGCAACGATTC 92
   |||||
   |||||
16 AlaValProLysSerLeuAlaGlySerThrSerAsnIleProAsnGluLe 32
   |||||
93 TTTATCCGGCAGGTCTCGACCGTCAGCATTTGCAACCCGACGAGAAAT 142
   |||||
32 urPheVal.....GlyProAspIleArg 40
   |||||
143 ACCACCTATTGGCAGCAGGGGGGACCTTGGCCGACGCGATATC 192
   |||||
40 rOlSerGlnAsnAlaGlnIleValSerLysSerSerLysValVal 56
   |||||
193 GGATTGGGAACATACAAAGCCATCAGTTGGGCAACCTGTTCAATCACA 242
   |||||
57 ProGlnAlaLeuValProGluAspAspIle..... 66
   |||||
243 GCGCGCCATTAAAGGAAATATCGGCTAC.....ATTGTCGCT 280
   |||||
67 ArgIaValAlaGluGlyIleLeuAsnIleGlySerAsnAspAsnArg 83
   |||||
281 TTTCGGATCAGCGGCAC.....GAATCCATTCCTCCCTTGACAAAC 321
   |||||
83 rOlValSerHisThrIleThrPheValGluLeuGlnLysSerHisGlnAsn 99
   |||||
322 CATGCTCATTCGATTCCT.....GATGAAGC 350
   |||||
100 HisLeuThrGlnAsnAspArgAsnPheGlyThrSerArgLeuAspAspVal 116
   |||||
351 CGGTAGTCCCGTTCGACGATTCAGCTTTACCGCATTCATGGACGAT 400
   |||||
116 IAlaProAsnAlaAspGlyValArgArgLeuArg.ThrSerGlySer 132
   |||||
401 ACG.....AACACCATCCCGCGCAGCGCTATGACGGCCACAGGGCGC 444
   |||||
133 ThrGlyLeuSerAsnAlaProProSerAlaAsnValSerLysAlaSer 149
   |||||
445 GGCTATCCCGTCCCAAGGCGCAGGATATATACAGCTACGACATATAA 494
   |||||
149 rAsnLeuSerLeu.....AlaSerLeuAlaLysThrGlnProGlu 163
   |||||
495 AGCGCTGGCCCAAAATATCCGCTCACTGACGCAACAGCAGCA... 541
   |||||
163 rGAlaThrProGluValCysValProLeuAsnProAspThrGlySerVal 179
   |||||
542 .....CCGACAAAGGCTTGTG..... 558
   |||||
180 ProLeuIleHisProGluGlnThrAspArgGlyLeuProTyrAlaProAs 196
   |||||
559 .GACCTTTCCCAATACCGGTATATG...CTGACGCAAGAGAGAGG... 603
   |||||
196 rGluLysPheHisAsnSerGlySerLeuLysLeuProLysGlyAlaSerL 213
   |||||
604 ..GACGATTCAAACGCGCCACCGCATACGCCCGAG.....CTG 642
   |||||
213 euGlnAspLeuSerArg.....SerProSerArgAlaValLeu 225
   |||||
643 GACAGATCGGCAATCGCCCGCAAGCT..... 669
   |||||
226 AsnGluAspIleValAsnValAspGluCysAlaProProGluProTyrL 242
   |||||
670 ....TTCACGCGCATCGAGATATCTGCAAAAC...ATCATCGGC... 708
   |||||

```

```

242 nGluTyrAsnGluValLeuAspAspValGluAsnAlaValAlaGlyThrS 259
   |||||
709 .....GCGGACAGCAAAATGTCGGC 729
   |||||
259 erProLeuGluTyrThrSerLysProLeuAlaAlaAsnAlaGlnArgSer 275
   |||||
730 GCAGCGATGCCGTCGACGAGTATAACGCAAGCGTCAAAATGCTGTAT 779
   |||||
276 ThrAlaAsp.....LeuThrGluSerAsnIleCys..... 286
   |||||
780 GCAGCGCTGGGCTGCTTCACCCGAAACACAGATGGCGCATCAACG 829
   |||||
287 .....GlyLeu.....ThrAlaGlyLysSerAspProValIle 298
   |||||
830 ATTTGGCAGATATGGCGACATCAACGACTATGGCGACGACCATCGC 879
   |||||
298 spVal...AspLeuSerGlnThrIleAspGluGlnSerIle..... 310
   |||||
880 GATTGGGCGATCCAAACCCCAATGCCGACACAGCATAGAACCCGTAG 929
   |||||
311 .....ProGluAlaGluLysGlyPheTyrThrLysAs 321
   |||||
930 CAATATCTTTAGCGCAGTCATCCCGTCAAAAGGATGGAGCTGTCCG 979
   |||||
321 rGlyGluGlyThrAlaGlyLeuProPheAsp.....IleValS 334
   |||||
980 GAAATACGCTTGGCGGATCAGCGCATCGCATCTGTCAAGCGTGGAG 1029
   |||||
334 erAsnLeuAspIleProAsnGluAsnAlaHisGluSerSerArgSerL 350
   |||||
1030 ATGGCGCAGATCGCATTCGCCGAAAGGAAATCCCGCTGACGACAAAT 1079
   |||||
351 LysLysHisThrGly...ProSerLeuSerSerAlaSerGlnProSerL 366
   |||||
1080 TGCGGATGGCGCATACGCCAAATACCCCTCCCTTACATTCGCCAAAT 1128
   |||||
366 aAlaSerSerSerSerSerSerLeuProSerAsnLeuAspLysIleAsn 383
   |||||
1129 ..ATCGTTCAAACTTGAG.....CAGGTTAC 1155
   |||||
383 spAlaLysLysAsnIleGluValSerAlaAsnGluProGlnProAlaPro 399
   |||||
1156 GCGAAGGAAACATCACTCCCTCA.....ACGCT 1184
   |||||
400 ValLysLysAspValProLysSerGlnValGlyGluThrAspThrH 416
   |||||
1185 GCGCGCTCAAGGAAAGATGTGAATGCAAACTGCAAAACAGCCACCGA 1234
   |||||
416 rAspValIleAsnAsnSerThrProLysGluGluThrGluGluSerPro 433
   |||||
1235 AGACCAAAAGTCCGTTGACGGAAGGTTTCCGAAATTTGAAAAAGAC 1284
   |||||
433 erThrGluLeuProGluThrGlyLysGluGlnProAsnLysAlaGlu 449
   |||||
1285 GTAATAATACGATACGAGATTAATACCGCTGTACA...CAAGTGAATCC 1331
   |||||
450 ProAlaValProThrGluAlaSerSerThrLysProSerGluAlaAla 466
   |||||
1332 TATATGAAACCCGCTTTAAT.....CCTAAAGTTCGTGGATCGG 1375
   |||||
466 uGluSerThrProArgPheSerValArgProAsnLysPheThrGlySer 483
   |||||
1376 CTCATTTCTGG.....TCTATAACTGCCAGAAATCAATACGCAAAATTA 1419
   |||||
483 rGAlaGlyPheValAlaAlaLeuGluSerArgLeuGlnLysGlyProLeu 499
   |||||
1420 CCAAGCGAAGGTAGATACATATATCCACCTAAATAATTAATCTCTCT 1469
   |||||
500 MetArgSer.....PheValProAsnLysSerLysSerProse 512
   |||||
1470 AGCAGCGCTACCAAAAGGACCTAATATGATATTTGGATAAATTTGTA 1519
   |||||
512 rGly.....ThrLysSerProAlaSerGlyGluThrSerGluAlaGly 527
   |||||

```

1520 ATGAATGGACTAAA 1533

527 allysglutrhglu 531

seq_name: p1r2:T31422

seq_documentation_block:

C-terminal domain-binding protein RA9 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 07-Dec-1999

C:Accession: T31422

R:Kuruyev, A.; Patwajan, M.; Litington, Y.; Joshi, R.V.; Gentile, C.; Gebara, M.; Cord

Proc. Natl. Acad. Sci. U.S.A. 93, 6975-6980, 1996

A:Title: The C-terminal domain of the largest subunit of RNA polymerase II interacts with

A:Reference number: 221024; MUID:96293459

A:Accession: T31422

A>Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 1-1473 <YUR>

A:Cross-references: EMBL:U49057; NID:g1438533; PID:g1438534; PIDN:AAC52658.1

A:Experimental source: hippocampus

alignment_scores: Quality: 137.00 Length: 624
Ratio: 0.566 Gaps: 33
Percent Similarity: 38.782 Percent Identity: 21.314

alignment_block:

US-09-303-518d-465 x T31422 ..

Align seg 1/1 to: T31422 from: 1 to: 1473

```
138 GAATATCCGATTCGCGAGAGGGGAGTCCGAGCGGACGCGTC 187
||||| ||| :||| |||||:||||| |||||
719 GlutIeGluprOthgluIleInIseSerAlaArgAlaGlnArg. 734
188 ATATCGGATTGGGAAATACAAAGCCATCAGTTGGGACACCTGTTCATC 237
||||| :||| |||||
725 .....ProSerProProAspProtrrAspa 743
238 CAGCAGGCGCCATTAAAGAAATATCGGCTACATTGT...CGCCTTTTC 284
||| :||| |||||
743 spgluAspGly.....ValserCysThrProPhehe 753
285 CGATCAGCGGACGAGATGCATTCGCCCTTGACACACGCTCACACT 334
||||| :||| :||| |||||
754 GlySerIgluIuArgThr.....ValThrCysValThr.. 764
335 CGGATTCTGATGAACCCGCTAGTCCGTTGACGATTCAGCCTTACCGC 384
764 .....
385 ATCCATTGGGACGATAGACACACCATCCGCCGAGCGCTATGACGGGCC 434
:||||| :||| |||||
765 .....ValIgluIuProSer..... 769
435 ACAGGCGCGGCTATCCGCTCCCAAAAGCGCGAGGATATATACAGCT 484
769 .....
485 ACGACATAAAGCGCGTTCGCCAAATATCCGCTCAACCTGACGACGAC 534
||||| :||| |||||
770 .....ValProSerProAsp...Alap 776
535 CGCAG.....CACGAGACAGCGCTTCGACCGCTTCCACAAATAC 575
||||| ||||| :||| |||||
776 rogluIleThrThrIleArgIleValIgluPheArgAlaSerSerArgSer 792
CGGTAGTATGCTGACGACGAGAGTAGGCGAGGATTCAAACGCCGCCACCC 625
||| :||||| :||| |||||
793 ArgSerThrSerSerArgSerArgIgluIuIuIuIuIuIuIuIuIuIuIu 809
```

```
626 GATA.....CAGCCCCGAGTGCAGAGA 648
:||||| :||| |||||:|||||
809 sValAlaArgGluHisIleArgThrArgSerSerThrArgSerClySerA 826
649 TCGGCAATGCCGCCGAGACTTTCACGCGACATGCAATATCGTCAAAA 698
:||||| :||| |||||
826 rgaSparArgThrSerArgSerValSerProPheThrGluHisThrIlys 842
699 CATATCGGCGCGCGCAGAAATATGTCGGCGGACGCGATGCCGTGACG 748
:||||| :||| |||||
843 ArgHisArgAlaIuThrIuSerArgArgSerSerSerArgAlaIse 859
749 GTATACGCAAGCTCAACATCTCTTATGCAGCGCTGCTGCTGCTT 798
:||||| :||| |||||
859 rserGluAspArgAlaIuArg..... 866
799 TCCACCGAAACAGATGGCGCATCAGATTCGACAGATATGCGCGCA 848
:||||| :||| |||||
867 ..ArgIuAspArgAspArgGluHisArgArg...GlyProThrIly... 880
849 ACTCAAGACTATGCCGACACCATCCGCGATTGGGAGCTCCAAAC 898
||| :||||| :||| |||||
881 HisGlyArgCysThrArgIuSerArgSerArgSerClySerProIlyse 897
899 CCAATGCCGCAAGCGATGAGCGCGTCAAGATATCTT.....TACG 942
:||||| :||| |||||
897 rSerSerCysGluArgHisIuSerArgArgArgIuIuIuIuIuIuIuIu 914
943 GCAGTCATCCCGTCAAGGAGATTCGCTGTTGGGGAATATACGCGCTT 992
:||||| :||| |||||
914 LyserArgSerArgIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIu 930
993 GGGCGCATCAGCGCATTCCTGCAAG..... 1020
931 AspArgArgHis..LysHisArgIuIuIuIuIuIuIuIuIuIuIuIuIu 947
1021 .....CGTCCGAGATG..... 1033
947 ysgIuSerMetThrArgSerArgGluArg..ArgArgITrArgSerArgse 963
1034 .....GCGAGATCGCATTCGCGAAGGAAATC 1061
963 rProSerValIgluHisArgThrArgArgProHisSerArgIuIuIuHis 980
1062 GCGCGTCA.....GCGACATTTTGGCGATGGCGGCTACGCGCA 1099
||| :||| :||| |||||
980 roHisSerProGluIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIu 996
1100 .....AATACCGTCC... 1111
997 AlaprogIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIu 1013
1112 .....CTTACATTCGCGAATATTCGCTCAAACTTGAGCA 1148
:||||| :||| |||||
1013 oValSerGluValSerValIuIuIuIuIuIuIuIuIuIuIuIuIuIu 1029
1149 GCGTTCAGCGCAAGAAACATCACTCTCAACCGCGCGCGTCAACG 1198
:||||| :||| |||||
1029 IuValIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIu 1045
1199 GAAAGATGTGAACCTGCGCAACCAAGCGCGCGGAGCAAGAGTCCG 1248
:||||| :||| |||||
1045 IagIuSerValSerCys.....ValPro 1052
1249 TTTGAC.....GTTAAAGGTTTCCGAA 1271
||| :||| :||| |||||
1053 GluAspIuAspIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIuIu 1069
1272 TTTTGAAGAAAGAC.....GTAAATATGATAGAGAAATTAATACG 1312
:||||| :||| |||||
1069 pHisSerAsnGluAlaIlePheIleIuIuIuIuIuIuIuIuIuIuIuIu 1086
```



```

172 .....ProSerSerProAlaProProLeuProLeuSerg 183
626 ..GATACAGCCCGAGCTGG...ACAGATCGGGCAATGCCCGCAAGCT 670
183 lylaserlaleProlyValProGlnasnArgProHisMetProSerVal 199
671 TCACGCGCACTGCAGTATGTCMAAAACATCATCGCGCGCGAGAGAGA 720
200 ArgProAla.....HisArgSerHisGlnArgly 209
721 ATGTGCGCGCAGCGGATGCCGTGAGGATATAGCGAAGGCTCAAAACAT 770
209 sserSerAsnIleSerLeuProSer..... 217
771 TGCTGTATGACAGCGTTGGTCTGCTTCCACCGAAACAGATGGCG 820
218 .....ValSerAlaProProLeuPro..... 224
821 GCATCAACGATTGGCAGATATGGCGCAACTCAAGACTATGCCGAGCA 870
225 .....SerAlaSerLeuProThrHi 231
871 GCCATCCGCGATTGGGCACTCCAAACCCCAATGCCG..... 907
231 sValSerAsnProProGlnAlaProProProProProProThrIleG 248
908 .....CACAAAGCATAGAACCG.....TCAGCATATCTTTACGG 943
248 lylLeuaspSerIysAsnIleLysProThAspAsnAlaValSerProPro 264
944 CAGTCATCC...CCGTCAAGGAGATTGAGCTGTTCGGGGAATAAAGCGC 990
265 SerSerGlnValProIleGlyLeuProPheLeuAlaGlnIleAsnAl 281
991 TTGGCGGCATCAGGCGACATC.....CTGTCAACGGTCCGACAT 1031
281 aArgArgSerGlnArgIleValAlaGlnIleValSerSerThrLysIleG 298
1032 GGGCGAGATCGCATGCGCAAGGGAATCCGCG.....TCAGCG 1072
298 lInThrIleAsnHisLysSerProSerGlnProProLeuProSerSerAla 314
1073 ACAATTTTCCGATGGCGCATAGCGCAATACCCGTCCTTCATTCATCC 1122
315 ProProIleProThrSerHisAlaProProLeuProProThrAlaProPr 331
1123 CGAAATATCCGTTCAACTTGGAGCGGTACGGCAAAAGAAACATCA 1171
331 oProProSerLeuProAsnValThrSerAlaProLysAlaThrSerA 348
1172 .....CCACTCAACGTCGCGCGCGCTCAAAAGGAAAGA 1204
348 lalProAlaProProProProProProLeuProAlaAlaMetSerSerAlaSe 364
1205 ATGTGAAGACTGGCAAAACAGCCA...CCCGAAGACCAAGTCCGCTT 1251
364 rThAnsnSerValIysAlaThrProValProProThrLeuAlaPro... 379
1252 GACGGTAAAGGTTTCCGAATTTTGAAGAAAGCATTAATACATACGAG 1301
380 .....ProLeuProAsn..... 383
1302 AATTAATACCGCTGTACCA.....CAAGTGAATCCTATGAGAG 1339
384 .....ThrThrSerValProProAsnLysAlaSerSerMetProAlaPro 399
1340 AACCCGCTTTAATCCTAAAGTTCTGTGGATCGGCTATTCTTGCTCT 1389
399 roProProProProProGlnAlaPheSerThrSerSerAlaLeu 415
1390 ATTAAGTCGCAATTCATACGCAAAATTACCAAGGCAAGGTGATCAG 1439

```

```

416 SerAlaSerSerIleProLeuAlaProLeuProPro..... 428
1440 ATATATCCACCTAAATAATTACTCT.....CCTTCAGACCGCTAC 1480
429 .....ProProSerSerValAlaThrSerSerValProSerAlaProPro 443
1481 CA 1482
443 ro 443
seq_name: p1r2:T42730

```

```

seq_documentation_block:
Bassoon protein - mouse
C:Species: Mus musculus (house mouse)
C:date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42730
R:Dieck, S.; Sammarti-Vila, L.; Langnaese, K.; Richter, K.; Kindler, S.; Soyke, A.; W
J. Cell Biol. 142, 499-509, 1998
A:title: Bassoon, a novel zinc-finger CAG/Glutamine-repeat protein selectively local
A:reference number: 222249; MUID:98345363
A:Accession: T42730
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3942 <DIE>
A:Cross-references: EMBL:Y17034; NID:G3413809; PIDN:CAA76598.1; PID:G3413810
A:Experimental source: strain 129 SVJ
C:Genetics:
A:Map position: 9F1
A:Introns: 72/2; 208/3; 505/3; 675/3; 2889/3; 3582/1; 3851/3; 3886/1; 3930/1
A:Note: bassoon
C:Function:
A:Description: may be involved in cytomatrix organization at the site of neurotransmi
A:Note: component of the presynaptic cytoskeleton
C:Keywords: coiled coil; zinc finger

```

```

alignment_scores:
Quality: 135.50 Length: 541
Ratio: 0.560 Gaps: 27
Percent Similarity: 44.732 Percent Identity: 22.551

```

```

alignment_block:
US-09-303-518D-465 x T42730 ..

```

```

Align seg 1/1 to: T42730 from: 1 to: 3942

```

```

95 TTATCCGCGAGCTTCTGACCGTCAGCATTTTCGAACCCGAGCAATAC 144
||||| : : : : : ||| : : : : :
306 leuSerThrLysProSerThrAlaGlnProArgProProAlaGlyAla 322
145 CACATATTCGCGACGACGAGGGGAACTTCCGAGCGACGCGTCATTCAG 194
322 aglnIlysserAla...ThrThrValProSerGlyLeuGlyAlaGly 338
195 ATGTGGAACATATACAGCATCAGTTGGGCAACCTGTTTCATCCAGAGG 244
: : : : : : : : : : : : : : : : : : : : : : : : : : :
338 lu...GlnThrGlnGlnIlyLeuThrGlyLysLeuPheGlyLeuGlyAla 353
245 CGGCATTAAGGAATATTCGGCTACATTCGCCGTTTCCGATCAGCAGG 294
||| : : : : : ||| : : : : : ||| : : : : :
354 SerLeuLeuThrGlnAlaSerThrLeuMetSerValGlnProGlnAlaAs 370
295 CACGAAGTCATTCGCCCTTCGACAAACCATCCATTCGATTCGATTCGA 344
||||| : : : : : ||| : : : : :
370 pThrGlnGlnProProSerProSerLysGlyPro..... 381
345 TGAACCGCGATGTCCTGAGCGATTGACGCTTACCGCA..... 385
||| : : : : : ||| : : : : : ||| : : : : :
382 .....ProLysIleValPheSerAspAlaSerLysGlnAlaGlyProArg 396
366 .....TCATTTGGGAGGATACGACACCATTCGCG..... 415

```

```

397 ProProGlySerClyProGlyProGlyProThrProGlyAlaIalysThrGI 413
416 .CGACGGCTATGACGGCCACAGGGCGGCTATCCGCTCCCAAG 464
413 uProGlyAlaIalysThrGIProGlySerClyProGlyAlaIalysThr 430
465 CGGAGGAGTATATACAGCTACGACATAAAGCGCTGCCCAAAATATCC 514
430 hteIyGly.....GluHISGlnAlaIalysThrSAla 432
515 GCCTCAACCTGACCGACACGACCGGACCAAGCGCTGTCGACGCT 564
433 .....ThrAlaSerProIySHISGlyAla..... 441
565 TTCACAAATACCGGATAGTATGCTGACGCAAGAGTAGGCGAGGATTC 614
442 .....GluHISGlnAlaIalysThrSAla 450
615 ACGCGCCACCGATACAGCCCGGAGCTGACAGATCGGCAATGCCG... 661
450 lAlaIalysProIySHISGlyAla.....GluHISGlnAlaIalys 464
662 CGGAACCTTTCAACGGCACTGCAGATATGTCAAAATCATCTGCGCGG 711
465 ProIeuCysGlnAlaIalysnValGlySerArg...GlyProAlaIas 480
712 GGAGGAGAAATTTGCGCGCGAGCGATGCCCTGACAGGATTAAGCAAG 761
480 nTyraSHISGlyAla.....CysIySL 489
762 CTCAAACATTCGCTTATGC..... 781
489 euGlnValCysasnIeuCysGlyPheasnProThrProHISIeuValGlu 505
782 ...ACGGCTTGGCTGCTTTCACCGCAAAACA...AGATGGCGGCGATC 825
506 lysThrGIuThrIeuCysIeuasnCysGlnThrIySHISGlyIeuGluGI 522
826 AACGATTTGGCAGATATGCGCGCACTCAAAAGATATGCCGACGACCAT 875
522 ySerIeuGlyGIuProAlaProIeuProIeuProThrProGlnGlnProp 539
876 CGGCAATTTGGCAGATTCACAAACCCCAATGCCGACAAAGCATAGAAGCG 925
539 roAla.....GlyValProHISArgAla.....Ala 547
926 TCAGCAATTTCTTACGCGATCATCCCGTCAAGAGGATTCGACGCTGT 975
548 GlYAlaIalProIeuIySHISGlyIeuGlyIeuGlyIeuGlyIeuProse 564
976 CGGGGAAAATACGCTTGGCGGCGATCAAGGACATCTGTCAACGCGTTC 1025
564 rIySerIeuProAlaIalysAlaSerProGlnAlaThrIySHISGlyAla 579
1026 GCAGATGGGCGAGATCGCATGCCGAAGGAAATCCCGCTGACGCGACA 1075
580 .....ProGlnAlaThrIySHISGlyAlaSerProGlnAlaThrIySHIS 591
1076 ATTTTCCGATG.....CGGATACGCGCAATACCGCTCCCTACCAT 1119
592 AlaserProGlnThrIySHISGlyAlaSerProGlnAlaIalysProIeuArgAl 608
1120 TCCGCAAAATATCCGTTCAAACTTGAGCAGCGGTTACGCAAGAAACAT 1169
608 aThrGIuProSer...LysThrSerSerAlaGlnGluIySHISGlyIeu 624
1170 CACCTTCCTCAACCGTGGCGCGGTCAAGGAAAGATGTGAACCTGGCA 1219
624 lathProAla.LysAla.....GluProValPr 633
1220 ACAAAAGCCACCGCAAGCAAGATGCGCTTGAACGTAAGGCTTTCG 1269
633 OlySHISGlyIeuProGlnThrIySHISGlyAlaSerProGlnIySHIS 646

```

```

1270 AATTGTAAGAAAGAGCTAAATACGATACGAGATTAATACCGCTGACC 1319
647 LysAlalysSerGlyValIys.....ArgThrAspProAlaThrPr 660
1320 ACAAGTGAATCTTATAGATGACACCCGCTTTAATCTTAAGAGTTGTGTCG 1369
660 oValIalysProValProGlnAla.....ProIySHISGlyIeuGlu 674
1370 GA.....TCGGCT 1377
674 lAGluGIuProValProIySHISGlyAlaSerGlnAspIeuSerAlaSerPro 690
1378 CATCTTGGCTATTAATGTCGACGATTAATCAATACGCAAAATTAACAGCA 1427
691 GlnSerIeuSerAspThrGlyIySHISGlyAlaSerSerAlaSerSerGly 707
1428 AGGTAGATCAATATATCCACCTAAATTAATAC.....TCTC 1465
707 nSerGluIleThrGlyAlaIalysGlnGluIySHISGlyAlaSer 724
1466 CTTCAGCAGCGCTACCAAGAGACCTAATATGATATTTGATTAATTT 1515
724 lAGlyValIThrGlyProIArgProIArgProIArgProIArgProIArgPro 740
1516 GGTAAAT..... 1521
741 GlySerSerIeuAspProSerIeuGluIalGlnAlaValAlaProSerAl 757
1522 GAATGAGCTAAAGCTCA 1539
757 aGluThrSerIySHISGlyPro 763
seq_name: pIrf2:C39135
seq_documentation_block:
  hypothetical protein 3 (gyrB region) - Haloferax sp.
  C:Species: Haloferax sp.
  C:Date: 11-Feb-1993 #sequence_revision 11-Feb-1993 #text_change 18-Nov-1994
  C:Accession: C39135
  R:Holmes, M.L.; Dwyer, S.M.; Smith, M.L.
  J. Bacteriol. 173, 642-648, 1991
  A:Title: Mutations in DNA gyrase result in novobiocin resistance in halophilic archae
  A:Reference number: A39135; MUID:91100352
  A:Accession: C39135
  A:Status: preliminary; translation not shown
  A:Molecule type: DNA
  A:Residues: 1-437 <HOL>
  A:Cross-references: GB:M38373

```

```

alignment_scores:
  Quality: 135.00      Length: 466
  Ratio: 0.689        Gaps: 27
  Percent Similarity: 42.060      Percent Identity: 25.107

```

```

alignment_block:
  US-09-303-518D-465 x C39135 ..

```

```

Align seg 1/1 to: C39135 from: 1 to: 437

```

```

276 CGCGTTTCCGATCAGGGGACGAGTCCATTCCTCCCTGACAAACATG 325
21 ProIArgThrArgAlaGlnHISArgAspAspIeuProIeu.....Le 34
326 CCGTCAATTCGATTCGATGAAAGCGGTAAGTCCGCTTACGATTCAC 375
34 uAlaGlyArgArgArg.....LeuArgAspAspArgValArgIeuGln 49
376 CTTTACCGATCCATTCGAGCAGGATACGAACACATCCCGCGAGGCTA 425
49 sPAlaIArgIySHISGlyProIArgAlaIalysValProGlnIeuArg..... 63

```

426 TCAGCGGGCCAGAGCGG.....CGCTATCCGGCTCCCAAGGCGGA 469
 64GlyArgSerPheAlaLeuArgIleArgIleAspArgVal 76
 470 GGGATATATACAGCTACAGCATAAAGCGCTTCCCAAAATATCCGCTC 519
 76GlyIleValProLeuArgIleArgIleProArgValProG 93
 520 AACCTGACCGCAACCCGACGCGACAGCGGCTTCCGACCGTTTCCA 569
 93LeuArgSerPheAlaLeuArgIleArgIleAspArgVal 96
 570 CAATACCGGTATATCTACGACGAGAGTAGCGGACCGATTAACGCG 619
 97AspGlyLeuProArgIleArgIleArgIleArgIleArg 103
 620 CCACCGGATACAGCGCGA...GCTGACAGATGCGGACATGCGCGCA 666
 103GlyIleValProLeuArgIleArgIleArgIleArgIle 120
 667 GCTTACGCGGACCTG.....AGATATCGTCAAAACAT 701
 120GlyIleValProLeuArgIleArgIleArgIleArgIle 136
 702 ...CATGCGCGCGAGAGAAATTTGCGGCGACGCGATGCGTCCAGG 748
 137GlyIleValProArgIleArgIleArgIleArgIleArg 143
 749 GTATAGCGAGGCTCAAAACATGCTGTATGACGCGCTTGGTCTGTT 798
 144Ala 145
 799 TCACCGGAAACAGATGCGCGCATCAAGATTTGCGAGATATGCGCA 848
 145GlyIleValProLeuArgIleArgIleArgIleArgIle 161
 849 ACTCAAGACTATG.....CGACGACCGATCCGCGATGCGG 886
 162GlyIleValProArgIleArgIleArgIleArgIleArg 178
 887 C.....AGTCCAAACCC 900
 178ArgSerPheAlaLeuArgIleArgIleArgIleArgIle 195
 901 AATGCGCG.....ACAAG 914
 195LeuArgIleProArgIleArgIleArgIleArgIleArg 211
 915 CATGACGCGCTACGCAATATCTTTACGCGACGATCCCGTCAAGGGA 964
 212HisProProthrla 222
 965 TTGAGGCTTGGGGAATACGGCTTGGGCGATCAGCGGACATCT 1014
 222GlyIleValProArgIleArgIleArgIleArgIleArg 237
 1015 GTCAAGCGGTCAGATGCGGAGATGCGATTCGCAAGGAAATCCGC 1064
 238GlyIleValProArgIleArgIleArgIleArgIleArg 252
 1065CGTACGCGCAATTTTGGCGATGCGG 1090
 253ArgIleValProArgIleArgIleArgIleArgIleArg 269
 1091 C...ATAGCGCAATATCCGCTTCCGATTCGCAAAATATCCGTTCA 1137
 269GlyIleValProArgIleArgIleArgIleArgIleArg 286
 1138 AACTTGACGAGCTTACGCGCAAGAAACATCAACCTTCAACGCTCC 1187
 286GlyIleValProArgIleArgIleArgIleArgIleArg 302
 1188 GCGCTC.....AAACGAAAGATGTGAATCTGCAACAA 1222

303 GluValProGlyAspPheAlaProGlnGlyIleAspSerGluArgG 319
 1223 AACGCGCCGAGACGCAAGCTCCGTTGACGCGTAAAGGTTTCCGAT 1272
 319GlyIleValProArgIleArgIleArgIleArgIleArg 336
 1273 TT.....TGAAAGAGCTAAATACGATACGATTAATATAC 1310
 336ArgIleValProArgIleArgIleArgIleArgIleArg 351
 1311 CGCTGACCAAGAGTAATCTATAGTAGAACCGCTTATATCTTAAG 1360
 351GlyIleValProArgIleArgIleArgIleArgIleArg 368
 1361 GTTCTGCGGATGCGGCTCATTTGCTGTATACGCGCAATCAATAC 1410
 368AlaArg 375
 1411 GCAAAATTACCA.....AGGCAAGTAGATCAGATATAT 1445
 376AlaSerValProAlaHisAlaProAlaHisArgGlyArgLeu 391
 1446 CCCACCTAAATATCTCTCTTACGACCGCTTACCAAA 1485
 391ArgGlySerThrAlaValAlaProArgProLeuProArg 404
 seq_name: p1r2:G01763
 seq_documentation_block:
 atrophin-1 - human
 C:Species: Homo sapiens (man)
 C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
 C:Accession: G01763
 R:Margolis, R.L.
 submitted to the EMBL Data Library, March 1995
 A:Reference number: G08343
 A:Accession: G01763
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1184 <MAP>
 A:Cross-references: EMBL:U23851; NID:915325; PID:915326
 C:Genetics:
 A:Gene: GDB:DRPLA; B37
 A:Cross-references: GDB:270336; OMIM:125370
 A:Map position: 12p-12p
 alignment_scores:
 Quality: 134.00 Length: 682
 Ratio: 0.515 Gaps: 35
 Percent Similarity: 38.123 Percent Identity: 22.141
 alignment_block:
 US-09-303-518D-465 x G01763 ..
 Align seg 1/1 to: G01763 from: 1 to: 1184
 65 CACGCGCTCAGATTTGCAACGATCTTTATCCGCGAGGTTCTCGAC 114
 252 HisProPro.....ProthrlProIleSerValSerSerG 265
 115 CGTACGATTTGCAACCGGACGCAATACCACTATTCGCGAGGAGG 164
 265GlyIleValProProthrlProProthrlProValGly 282
 165 GGAACCTTGGCGAGCGGCTCATATCGATTTGGCAACATCAAGCC 214
 282GlyIleValProProProProAlaAspHeProHisValThr 298
 215 ATCAGTTGGGCAACCTTTCATCCAGCAGG..CGGCGATTAAGAAATAT 263
 298OAsnLeuProPro.....ProAlaLeuArgProLeuAsnAla 313

828 uleuGlunrGserValLysLeuAlaGlnGluGlyArg 840
 seq_name: plr2:T33369

seq_documentation_block:
 hypothetical protein H02F09.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_rev: 1000 29-Oct-1999 #text_change 17-Mar-2000
 C:Accession: T33369
 R:Geisler, C.; Harmon, G.
 Submitted to the EMBL Data Library, July 1998
 A:Description: The sequence of C. elegans cosmid H02F09.
 A:Reference number: Z21330
 A:Accession: T33369
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1275 <GEI>
 A:Cross-references: EMBL:AF077538; PUD:AC64622.1; GSPDB:GN00028; CESP:H02F09.3
 C:Genetics:
 A:Gene: CESP:H02F09.3
 A:Map position: X
 A:Introns: 42/1; 78/3; 106/3; 135/2; 161/2; 224/1
 C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

alignment_scores:
 Quality: 136.00 Length: 458
 Ratio: 0.642 Gaps: 20
 Percent Similarity: 46.288 Percent Identity: 22.052

alignment_block:
 US-09-303-518D-465 x T33369 ..

Align seq 1/1 to: T33369 from: 1 to: 1275

```

92 CTTTATTCGCGAGGTTCTCGACGTCGACGATTTCG...AACCGACGGG 138
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
177 ValIleThrIleThrIleValIleValSerLeuSerLeuSerIleThrAs 193
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
139 A...AATACCACTATTCGCGACGAGGGGGAATTGCCGACGACGG 185
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
193 pmetAsnSerArgTyrGlyLysAlaIleAsnIleProThr 208
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
186 TCATATCGGATTGGGAACATACAAAGCATCATGTTGGCAACTGTTC 235,
208 ..... 208
236 TCCAGACGAGCGCCATTAAAGAAATATCGGCTACATGTCCGCTTTC 285
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
209 .....ProthrGlnSplIleSerAsnLysIleAsnIleLe 221
286 GATCAGCGGACGAGTCCATTCCTCCCTTCGACAC 322
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
221 uasnlIleGlyThrThrGlnThrProValThrThrSerThrMetAla 238
323 .....ATGCCCTCAC.....ATTCGATT 340
238 hrThrThrIleAsnValThrSerAlaIleProAsnThrThrValThrIle 254
341 CTGATGAGACCGGATGCCCTTGACGATTCAGCTTACGACATCCAT 390
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
255 SerThrSerProThrThrValThrValProSerThrIleAlaGlnThrSe 271
391 TGGGACGATACGACACCATCCCGCAGCGGCTATGACGGCCACAGG 440
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
271 rSerThrThrThrValThrValProThrThrValThrGly...Prot 287
441 CGGCGGCTATCCGCTCCCAAGGCGGACGATATATACGACGACGCA 490
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
287 hrThrValThrValProThr.....ThrValValThr 298
```

```

491 TAAAGCGTTCCCAAAATATCCGCTCACTGACGCAAC... 535
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
299 IleProSerThrValThrSerProIleThrThrProSerThrValVa 315
536 .GCAGACCGGACACGCGCTTTCGACCGCTTCCAAATACGGTGAT 584
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
315 lThrValPro.....SerThrValThrValProSerThrVa 328
585 GCTGACGCAAGATGAGCGACGATTCAAACGCGCCACCGCAT 628
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
328 lValThrLysPro...SerThrValThrValThrProSerThrVal 343
629 .....ACAGCCCGAGCTGACGATCGGCAAT 657
344 ThrValProSerThrValThrLysProAsnThrValThrSerSe 360
658 GCCCGCAGACCTTTCACGCGCATGATGCAAAACATCATCGG 707
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
360 rProThrValAlaThrThrProThrThrValThrThrProSerThrV 377
708 CGGCGCAGAGAAATGTCGCGCAGCGATCCGTCAGGCTATAAGCG 757
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
377 alValThrValProSerThrValThrValProThrThrValThr 393
758 AAGGCTCAACATTCGCTGATGACGCGCTGCTGCTTCCACCGCA 807
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
394 AsnProSerThrValThrAlaProSerThrThrValThrValProth 410
808 AACAGATGCGCGCA...TCACGATTTGGCAGATATGCGCACA 854
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
410 rThrValMetThrSerArgSerThrValIleThrThrProThrThrGly 427
855 AGACTATGCGCGAG..... 868
427 lYSerSerProSerThrAlaGlyThrSerLeuAlaSerThrAlaValThr 443
869 CAGCATCCGCGATTTGGCAGTCCAAACCCCAATGCGCACAAGCAT 918
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
444 ThrGlnThrSerIleGlySerSerSerThrProLeuProSerGlnSerTh 460
919 GAAGCGTCAGCAATATCTTACGCGACTATCCCGCTCAAGGATTTGG 968
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
460 rSerLeuSerMetSerSerLeuSerThrThrThrProSer 473
969 AGCTGTTCGGGAAATATAGGCTGGCGGATACGACGCAATC..... 1012
474 .....SerSerThrAlaGlyAlaThrSerProAlaThrGlnGln 486
1013 .....CTGTCAAGCGGTGCGACGATGG.. 1033
487 SerThrLysProThrIleGlyThrSerMetSerSerGlyProThrThrVa 503
1034 .....GCGAGATGCGATTGCCGAAAGGAAATCCGCCGT 1067
503 lAlaProGlyAlaSerThrGlnSerThrValLeuGlnSerSerThrPros 520
1068 CAGGACAAATTTGCCGATG.....CGCATACGCCCAAT 1102
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
520 erglYThrThrValThrLeuProSerGlySerSerThrAlaThrAlaGly 536
1103 ACCGCTGCC.....CTTACGATTCGCCGAAATATCCGTTCA 1137
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
537 ThrSerProGlnAlaSerThrValThrThrValThrAspIleSerThrVa 553
1138 AACTTGAGCAGCGGTTACGGCAAA.....GAAACATACCTCTCTCA 1180
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
553 lSerGlySerThrValThrSerGlnThrAlaGlnSerSerLeuSerThrG 570
1181 CGGTGCCCGCTCAACGAAAGATGTGAACCTGCAAC.....AAA 1224
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
570 lSerProThrSerAlaGlySerSerIleSerThrValSerThrValSer 586
1225 CGCACCAGCAAGACCAAGTCCG 1248
```

seq_name: p1r1:EDBEIF
587 SerGlnProSerThrTyrIlePro 594

seq_documentation_block:

Immediate-early protein IE180 - suid herpesvirus 1 (strain Indiana-Punkhauser)
C:Species: suid herpesvirus 1
C:Date: 30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change 16-Feb-1997
C:Accession: S04713
R:Cheung, A. K.
Nucleic Acids Res. 17, 4637-4646, 1989
A:Title: DNA nucleotide sequence analysis of the immediate-early gene of pseudorabies virus
A:Reference number: S04713; M01D:89315207
A:Accession: S04713
A:Molecule type: DNA
A:Residues: 1-1460 <CHE>
C:Superfamily: herpesvirus immediate-early protein IE175
C:Keywords: DNA binding; early protein; transcription regulation

alignment_scores:
Quality: 134.00 Length: 607
Ratio: 0.575 Gaps: 30
Percent Similarity: 38.386 Percent Identity: 22.570

alignment_block:

us-09-303-518d-465 x EDBEIF

Align seg 1/1 to: EDBEIF from: 1 to: 1460

```

117 TCAGCATTCGACACCGGCGGAAATACCACTATTCGCGACGAGGGGG 166
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
75 AAlaAlaGlyAlaThrArgProProArgProProSerAlaGlnGln.... 89
167 AACTGCCGAGCGACGCGTCATATCGATTGGG..... 200
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
90 .....GlnArgHisAlaArgArgGlySerGlyGluIleValV 102
201 .....AAACATACAAAGCCATTCAGTT 221
102 AlLeuAspSerGlnAspGlnGluAspGlnProGlySerProAlaAla 118
222 GGGCAACCTGTT..... 233
119 GlySerProValGlyLeuSerIleArgAlaProSerThrValThrSer 135
234 .....CATCCAGCAGGGCGCATTAAGGAAATATCGGC 267
135 rSerGlyProGlyProGlyProAlaProGlyProGlyArgArgProAl 152
268 TACATTGT.....CCGCTTTCCGATCAGGGCA 296
152 LmHisSerLmArgLmArgProGlyProProAlaAlaProGlyAlaArg 168
297 CGAAGTCACCTCCCTTCGACACACATGCTTCACATTCGATTCGATG 346
169 ProProGlnProProArgProPro..... 177
347 AAGCGGTGTCCTTCGACGATTCAGCTTACCGCATTCATGGGAC 396
178 .....ProProAlaProProAlaPro 185
397 GGATACGAACACCATCCGCCGACGG..... 422
185 rProAlaProProAlaProAlaArgArgArgGlyAspGlyProProArg 201
423 .....CTATGAGGGCGACAGGGCGGCTATCCCGCTCCCAAG 463
202 GlyGlyThrArgSerValSerProGlyArgArgGlyLeuGlyProArg 218
464 GCGCGAGGATATATACAGCTACGACATTAAGGCGT..... 500
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```

218 gArgHisGlnHisSerGlnGlnArgTyrProGlnArgArgHisGlyG 235
501 .....TGCCAAATATTCGCCCTCACTGA.....CCGACA 532
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
235 LyrProLeuProGlnProProProProGlyArgSerArgArgProAla 251
533 ACCGAGCAGCCGACACACGCTTCGACG...GTTCACAAATACCGGT 579
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
252 AlaAlaAlaProProProAlaGlnGlyThrAlaValValThrIle.... 266
580 AGTATGCTGACGCAAGAGTAGGACGATTCACACCGCCACCC.... 625
267 .....ThrSerThrLaserProTyrPl 274
626 .....GATACAGCCCGAGCTGACACA 646
274 euAspGluProAlaAlaAlaArgArgLeuAspProAlaAlaAlaTyr... 289
647 GATCGGCAATGCCGCCGAGCTTCACACGCGCATCGCATATGCTCAA 696
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
290 ...ArgProGluProArgLeuGlnProGlnLeuGlnLeu 305
697 AACATCATCGGCGGCGAG..... 716
305 LshHisArgArgArgArgAlaArgArgProArgProArgGluArg 321
717 .....AGAAATG 724
322 GlyArgThrArgProArgArgGlyArgGlyAlaProLeuGlnArgLm 338
725 TCGGCGCAGCGCATGCCGCGGATTAAGGAGCTCAAAATTCCT 774
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
338 oArgArgArgArgAlaGlyGlnGlyAlaLeuArgArg..... 350
775 GTTATGACAGGCTTGCTGCTTCACCGAAACAGATGCGCGCAT 824
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
351 .....GlyArgGlyPheSerSerSerSerSerSer 360
825 CAACGATTTGGCAGATATGCGCGCACTCAAAAGACT.....TGCGCAG 868
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
361 .....GlySerAspSerAspLeuSerProAlaArgSe 371
869 C...AGCCATCCGCGATTGGCGATCCAAACCCCAATGCCGACAGGC 915
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
371 rProSerAlaProArg.....AlaProAlaAlaAlaAlaAlaAla 386
916 ATAGAAGCCGTCAGCAATATCTTACGGCATGATCCGCTCAAGGAT 965
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
386 rArgSerAlaSerSerSerSerSerSerSerSerSerSerSer 402
966 TGGAGCTGTTGCGGGAATATACGCTTGGCGGATCAGCGCATCTG 1015
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
403 SerSerSerGlnGlnGlnAspGlnGlyValArgProGlyAla 419
1016 TCAAGCG.....TCGCAGATGGCGCGCATCGCATGCGCG 1050
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
419 euAlaAlaArgAlaGlyProProProSerProProAlaAlaAlaPro 435
1051 AAA...GGGAATCCGCCGACGACGACATTTTGCAGTCGCGCATACGC 1097
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
436 ArgProSerAlaSerSerAlaSerAlaThrSerSerAlaAlaAla 452
1148 AGCTTACGCGCAAGAAACATCACTCCCAACCGCGCGCTCAAC 1197
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
469 eThrAsnAsnHisLeuSerLeuMetAlaAspGlyProProPthAsp 485
1198 GGAAGAATGTGAACACTGGCAACCAACGCCCGCAAGCAAGTGC 1247
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
486 GlyProLeuLeuThrProLeuGlyGluProThrProGlySerAspPro 502
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```

1248 GTTTCAGCGTAAAGGTTTCGAAATTGTAAGAAACCTAAATAC... 1293
1249 | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
502 oAlaspllyarg.....Valarglyrglyg 511
1294 .....GATACGAGA.....ATTAAATAC 1311
511 lYAlaglyAspSerArglyluGlyLeuThrPaspGluAspValArgly 527
1312 GCTGACACACA.....GTGAATCTATAGATGAACCGGCTT 1349
528 AlaAlaAlaArglyArglyAlaAlaAlaGlyProVal.....ProValPh 542
1350 TAATCCGTAA..... 1359
542 eileProGluMetGlyAspSerArglyGlnHisGluAlaLeuValArgly 559
1360 .....GTTCTGTGCGATCGGCTCATCTGTGCTATACAGGCC 1398
559 euileTySerGlyAlaAlaGlyGluAlaMetSerTrpLeuGlnAspPro 575
1399 AGAATTCAATACGCAAAATTACCAAGCAAGTACAGATATATCCC 1448
576 ArgMetGln.....AlaProAspGlnArgPheasnGlnPheCysG 589
1449 ACCTAAAAATTACTCTCT..... 1467
589 naGArgValAlaHisAlaProHisGlnHisGlySerPheIleThrGlySerV 606
1468 ..TCAGCAGCGCTACCA 1482
606 alThrProProLeuPro 611

```

seq_name: p1r2:A53577

```

seq_documentation_block:
ascites stialoglycoprotein 1 - rat (fragments)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 12-Apr-1995 #sequence_revision 12-Apr-1995 #text_change 07-Feb-1997
C:Accession: A53577
R:Wu, K.; Fregien, N.; Carraway, K.L.
J. Biol. Chem. 269, 11950-11955, 1994
A:Title: Molecular cloning and sequencing of the mucin subunit of a heterodimeric, bifur
A:Reference number: A53577; MUID:94216302
A:Accession: A53577
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1630 <WUA>
A:Cross-references: GB:006746
C:Keywords: glycoprotein

```

alignment_scores:

Quality:	133.00	Length:	426
Ratio:	0.621	Gaps:	21
Percent Similarity:	50.235	Percent Identity:	22.535

alignment_block:

US-09-303-518D-465 x A53577 ..

Align seg 1/1 to: A53577 from: 1 to: 1630

```

92 CTTTATCCGCGAGGTTCTCGACCTGACATTTCGACCCGAGCGGAAA 141
734 MeclLeuThrThrAlaSerThrGluGlyThrSerGlyAspThrGlyHI 750
142 TACCACTATTGCGAGCGAGGGAAGTGGCGGCGGCGGCGGATAT 191
750 sThrThrAlaValThrThrGlnGlySerIleProAlaThrThrGlnLeus 767
192 CGGATTTG.....GAACATACAAAGCCATCAGTTGGCAACCGT 232
767 erThrThrPheAlaSerGlnLysMetSerThrValSerThrProThrThr 783

```

```

233 TCATCC.....ACAGCGCGCCATTAAG..... 256
784 SerSerIleGlnGlnLeuSerThrLeuProGlnSerGlnHisThrGlySe 800
257 .....GAATATCGGCTACATTTGCCGCTTTCCGATCAGCGGCGAGGT 302
800 rMetGluIleSerSerArgProGlnThrSerValThrSerThrLeus 817
303 CCATTCGCCCTTGACACACATGCTTCACATTCGATTCGTGTGAAGCG 352
817 erSerSerProSerGlySer.....ThrProValGlnThrArgSer 830
353 GTATGCCCGTTGACGATTCAGCTTACCGCATTCATCGGACGAGATAC 402
831 ValThrSerSerSerAspGlnArgThrAsnProThrSerSerGlyValSe 847
403 GAACACCATCCG.....CCGA 419
847 rAsnThrSerProAlaThrThrGluValLeuThrProThrSerSerProG 864
420 CGGCTATGACGGGCCACAGGCGCGGCTATGCCGCTCCCAAGCGCGA 469
864 lSerThrProGlyAsnThrAla.....ProArgThrThr 875
470 GGGATATATACAGCTACGACATAAAGCGTTGCCCAAAATATCCGCTC 519
876 GluThrSerThrThrThrThrThrThrThrThrThrThrThrThrThr 892
520 AACCTGACCGACA.....ACGCGACGCGGAGCAACAGGCT 554
892 nLysLeuProThrThrGlySerThrLeuGlyThrSerThrProThrGluValT 909
555 TGTCCACGCTTTCACAAATACCGTAGTACTGACGCAAGAGTAGCG 604
909 hThrThrLeuSerAlaSerSerSerAspGlnValGlnValGlnThrThr 925
605 ACGGATTCAAACGCCGCCACCGAT.....ACAGCCCGGAGCTGAC 645
926 SerGlnThrThrLeuSerProAspAlaThrThrThrSerHisAlaProAr 942
646 AGATGGGCAATGCCCGGAGCTTTCAGGCGACTGACGATATACGCA 695
942 gLysSerSerSerProProSerThrSerValIleLeuThrThrMetLAs 959
696 AAACATCATCGCGCGGCGAGGAGAAATGTCGGCGGAGCGATGCCGTC 745
959 erThrGlu..GlyThrSerGlyAsp.....ThrGlyHisThrThra 972
746 AGGATATAGCGAGGCTCAACATTCGTGTATGACAGGCTGGGTCTG 795
972 lValThrAspGlnGlySerThrProAlaThrThrGlu..IleSerVal 987
796 CTTTCACGGAACAGATGCGCGCATACAGATTTGGCAGATATAGC 845
988 ThrProSerThrGlnLysMetSerThrValSerThrLeuValThrSerTh 1004
846 GCAACTCAAGACT.....ATGCCGAGCAGGACCATCCGATTTGGGCG 889
1004 rGln..GluLeuThrSerSerGlnSerGlnArgThrGlySerMetGlyThr 1020
890 TCCAAACCCCAATGCCGCGACAGGCAATACAGCGCATATATCTTT 939
1021 SerSerLysPro.....GlnAlaThrThrProThrGluValAlaThrTh 1034
940 ACGGCAAG...TCATCCCGCTCAAGAGGATGAGCTGTGGGGAATAA 986
1034 rSerThrLeuSerSerSerSerArgly...SerLeuPheSer..... 1047
987 CGGCTTGGCGGATCACGGCACATCTCTGCAAGCGGTGCAAGATGGCG 1036
1048 .....Ala 1048

```


C:Species: Homo sapiens (man)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
 C:Accession: T47182
 R:Kocher, K.; Beyer, A.; Mewes, H.W.; Weil, B.; Wiemann, S.
 Submitted to the Protein Sequence Database, March 2000
 A:Reference number: Z24378
 A:Accession: T47182
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1299 <AAA>
 A:Cross-references: EMBL:AL162004
 A:Experimental source: adult testis; clone DKFZp434M1616
 C:Genetics:
 A:Note: DKFZp434M1616.1

alignment_scores:
 Quality: 132.50 Length: 567
 Ratio: 0.465 Gaps: 31
 Percent Similarity: 50.265 Percent Identity: 22.046

alignment_block:

US-09-303-518D-465 x T47182

Align seg 1/1 to: T47182 from: 1 to: 1299

```

74 CAGATTGG.....CAACGAT 90
   |||:||||
254 GluValThrPaspasnLysAlaAsnGluLysGlnArgSerGlnThr 270
   |||:|||||
91 TCCTTTATCCGGCGCTTCGACCGCAGCATTTCCACCCCGAGGAA 140
   ||| |||:|||||
270 rLysLeuProAspPheAlaLys.....LysGlnAlaThrGly 284
   ||| |||:|||||
141 ATACACCTTATCGCGACGAGGGGGAACCTGCCAGCGCGGCATCA 190
   |||:|||||
284 LeuGlnAlaGlnSerSerAlaSerValProProLeuAlaSerAlaPro 300
   |||:|||||
191 TCGGATGGGAACATATCAAGCCATCAGTTGGCAACCTGTTCAATCA 240
   |||:|||||
301 LeuProSerThrSerAlaSerValPro.....AlaSerThr 313
   |||:|||||
241 CAGCGCGCCATTAAGGAATATATCGCTACATTTCCGGCTTTCCGATCA 290
   |||:|||||
313 SerAlaProLeu.....ProAlaThrLeuThrProValProAla 327
   |||:|||||
291 CGGGCAGAGATCCATTCCTCCGACACCATGCTCCACATTT..... 334
   |||:|||||
327 ThrSerAlaProValProAlaSerThrLeuAlaProValLeuAlaSer 343
   |||:|||||
335 .....CCGATTCGTGATGAAGCGCGTAGCCGTTGACGATTCAGC 375
   |||:|||||
344 ThrSerAlaProValProAlaSerProLeuAlaProValSerAlaSerAl 360
   |||:|||||
376 C...TTTACCGCATTCATGGGAGGATAGCAACACCATCCCGCGACGG 422
   |||:|||||
360 AserValSerAlaSerValProAlaSerThrSerAlaAlaLeuThrS 377
   |||:|||||
423 CTATGACGGGCCACAGGGGGGCGATCCCTCCCAAGCGCGGAGGG 472
   |||:|||||
377 ThrSerSerProAlaSerAlaProAlaProThrProLeuAlaSer 393
   |||:|||||
473 ATATATACA..... 481
   |||:|||||
394 ValSerThrProAlaSerValThrLeuAlaSerAlaSerIleProI 410
   |||:|||||
482 .....GCTA 485
   |||:|||||
410 LeuAlaSerAlaLeuAlaSerThrSerAlaProThrProAlaProAla 427
   |||:|||||
486 CGACATTAAGGCGTTGCCAAATATCCGCTCAACCTGACCGGACGAC 535
   |||:|||||
427 LeuSerSerProAlaAlaProValIleThrAlaProThrIleProAlaSer 443

```

```

536 GCA.....GCACGGACACAGCGCTTGCACGG 564
   |||:|||||
444 AlaProThrAlaSerValProLeuAlaProAlaSerAlaSerAlaProAl 460
   |||:|||||
565 T...TCCCAATATCCGCTACTATGCTGACGCAAGAGATAGCGACGAT 611
   |||:|||||
460 AlaProAlaProThrProVal.....SerAlaProAla 470
   |||:|||||
612 CAAACGGCCGCCGATACAGCCCGAGCGACAGATTCGGCAATGCCG 661
   |||:|||||
470 snProAlaProPro.....AlaProAlaIleThrGlnAlaGlnThrHis 484
   |||:|||||
662 CCGAAGCTTCAACGGCAGCTGAGATATGCTCAAAAACATCATCGCGCG 711
   |||:|||||
485 LysProAlaGlnAsnProLeuGlnThrThrSerGlnSerSerLysGln 501
   |||:|||||
712 GCAGGAGAAATTG.....TCGGCGCAGGC...GATCCCGTGCAGC 748
   |||:|||||
501 opProProSerIleArgLeuProSerAlaGlnThrProAsnGlnThrAsp 518
   |||:|||||
749 GTATAAGCGAAGGCTCAACATTTGCTATGACAGCGTTGGCTGCT 798
   |||:|||||
518 yValAlaSerGlyLysSerIleGlnThrProGlnSerHisGlyThrLeu 534
   |||:|||||
799 TCCACCGGA.....ACAAGATGCGCGC.....ATCAGCA 830
   |||:|||||
535 ThrAlaGlnLeuThrPaspasnLysValAlaProProAlaValLeuAsn 551
   |||:|||||
831 TTTGGCAGATATGGCGCACTCAAGACTATGCCGACGAGCATTCGCCG 880
   |||:|||||
551 pLieserLysLysLeu.GlyProIleSerProProGlnProProSerVal 567
   |||:|||||
881 ATTGGCAGCTCCAAACCCCAATGCCGCAACAAGCATGAACCCGCTACG 930
   |||:|||||
568 SerAlaThrPaspasnLys...ProLeuThrSerPheGlySerAlaProSer 583
   |||:|||||
931 AATATCTTTACGCGCATCATCCCGCTCAAGGAGATTTGAGCTGTTCCGG 980
   |||:|||||
583 rGluGlnAlaLysAsnGlyGlnSer...GlyLeuIle...GlyT 598
   |||:|||||
981 AAAATACG.....GCTTGGCGGCGATCAGCGACATCCTGCA 1018
   |||:|||||
598 hrAspThrIleGlnPheGlnAlaProAlaSer..AsnGlyAsnGluAsnG 614
   |||:|||||
1019 AGCGGTC.....GCAGTGGCGGAGATTCGCTCCGCAAGG 1056
   |||:|||||
614 uValValProValLeuSerGluLysSerAlaAspLysIleProGluPro 631
   |||:|||||
1057 AAATCCGCGCTCAGGACATTTTCCGATCGGC.....ATACGC 1097
   |||:|||||
631 ySglnGlnArgGlnLysGln...ProAlaGlnGlyProIleLysAlaGln 646
   |||:|||||
1098 CAATATCCCGTCCCTTACCATTTCCCAATATTCCTGTAACACTGGAGC 1147
   |||:|||||
647 LysLeuProAspLeuSerProValGlnAsnLysGlnLysProGlyPro 663
   |||:|||||
1148 AGCGTTACGGCAAGAAACATCACTCTCTCAACCGTGGCGGCTCAAC 1197
   |||:|||||
663 oLleGlyLysGlnArg.....SerLeuLysA 672
   |||:|||||
1198 GGAAGAATGTGAACACTGGCAAAACGCCCGGAGCAAGT... 1244
   |||:|||||
672 snArgLysValLysAspAlaGlnValGluProGlnGlyGlnGluLys 688
   |||:|||||
1245 .....GCCGTTGACGTAAGGTTTCCGATTTTGAAGAACG 1285
   |||:|||||
689 ProSerProAlaThrValArg.SerThrAspProValThrThrLysGlu 705
   |||:|||||
1286 TAAATATCGATACGAGATTAATACCGCTGTACACAAAGTAATCTGTAA 1335
   |||:|||||
705 hrLysAlaValSerGlnMetSerThrGluIleGlyThrMetIleSerVal 721

```



```

1242 hrThrThrThrIle...ThrAsnGlyIleGlyValAlaAlaGlyThrThrAla 1257
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
988 GGCCTGGGGCGGCATCAGCGCATCTCTCATAGCGGTGCGAGATGGGCGA 1037
      :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1258 SerThrAlaValSerThrAlaSerThrThrThrSer..... 1270
1038 GATCGCATTCGCCAAGGAATCCGCCGTGACGACAAATTTGCCGATG 1087
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1271 .....SerGlyThrPheThrThrS 1277
1088 CGCGATACGCCAATAACCCGTCCTTACCATTC...CGAATATCCGT 1134
      :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1277 erCysThrSerThrThrThrThrThrThrSerSerIleSerAsnGlySer 1293
1135 TCAACCTTGAGCAGCGTTACGGCAAGAAACATCACCCTCCCAACCGT 1184
      :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1294 LysAspLeuProLysAlaMetIleLysProAsnValLeuThrHisValI 1310
1185 GCCGCCGTCAACGGAAGAATGTGAACGTGCACAAACAAACGCCACCGA 1234
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1310 e.....AspGlyPheIleIleGlnGlnAlaAsnGluProPhePro 1324
1235 AGACCAAGCGCGTTGACGTAAGGTTCCGAATTTGAAAAAGAC 1284
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1324 alThrGlnIleArgTyr.....AlaAspLysAsp 1333
1285 GTAATATGAGATGAGAAATTAATACCGCTGACACAAAGTGAATCCTAT 1334
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1334 ValSer..... 1335
1335 AGATGAACCCGCTTAAATCTTAA..... 1359
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1336 ..AspGluPro.....ProLysLysLysAlaThrMetGlnIleAspI 1349
1360 .....GTTCTGTGCGATGCGTCATCTGTGCTATACCTGCCAGA 1401
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1349 IeLysLeuSerGlyIleAlaSerAlaProGlySerAspMetValAlaCys 1365
1402 ATTCAATACGCAAAATTACCAAGCAAGTAGAATC.....AGATA 1442
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1366 GluGlnCysGlyLysMetGlnHisLysAlaLysLeuLysArgLysArgTyr 1382
1443 TATCCACACT 1452
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1382 rCysSerPro 1385

```

seq_name: p1r2:S29605

seq_documentation_block:

glycoprotein 350/220 - human herpesvirus 4
 C:Species: human herpesvirus 4, Epstein-Barr virus
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 26-Aug-1999
 C:Accession: S29605
 R:Klein, K.; Mueller-Lantzsch, N.
 submitted to the EMBL data library, October 1992
 A:Description: Sequences of the membrane proteins gp 350/220 and p140 of Epstein-Barr vi
 A:Reference number: S29605
 A:Accession: S29605
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-886 <KLE>
 A:Cross-references: EMBL:X67776; NID:959163; PIDN:CAA47966.1; PID:959164
 C:Superfamily: Epstein-Barr virus membrane antigen gp350
 C:Keywords: glycoprotein

alignment_scores:

Quality: 131.50 Length: 588
 Ratio: 0.502 Gaps: 28
 Percent Similarity: 44.558 Percent Identity: 22.279

alignment_block:

US-09-303-518d-465 x S29605 ..
 Align seg 1/1 to: S29605 from: 1 to: 886

```

110 TCGACCGTATGATTTCCGACCCGACGGAAATACCACCTATTCG..... 154
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
318 AsnThrThrAspIleThrTyrValGlyAspAsnAlaThrTyrSerValPr 334
155 .....GCAGCAGGGGGAGCTTCCGAGCGGACCGACGGTCATATCG 194
      :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
334 oMetValThrSerGluAspAlaAsnSerProAsnValThrAlaPr 351
195 ATTCG.....GAAACATACAAAGCCATCAGTTGGCA 226
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
351 heThrAlaThrProAsnAsnThrGluThrAspPheLysCysIleThrPhr 367
227 ACCGTGTCATCCAGCAGCGCGCATTAAGAAATATTCGCTACATGTCTC 276
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
368 LeuThrSerGlyThrProSerGlyCysGluAsnIleSerGlyAlaPheAl 384
277 CGCT.....TTCCGATCAGCGGC..... 295
      :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
384 aserAsnArgThrPheAspIleThrValSerGlyLeuGlyThrAlaPr 401
296 .....ACGAATCCATTCCTTCCGACACCATCCCTCAC 331
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
401 yThrLeuIleIleThrArgThrAlaThrAsnAlaThrThrThrHis 417
332 ATTCGATTCGATGAAAGCCGCTAGTCCCTTGACGATTCAGCCTTAC 381
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
418 LysValIlePheSerLys.....AlaProGluSerThrThrHis 430
382 CGCATCCATTGGAGCGATACGAACACCATCCCGCAGCGCTATGACG 431
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
430 rSerProThrLeuAsnThrThrGlyPheAlaAlaProAsnThrThrThr 447
432 GCCACAGGCGCGGTATCCGCTCCCAAGGCGGAGGATATATACA 481
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
447 Ly.....LeuProSerSerThrHisValProThr 456
482 GCTACGACATAAAGCGCTG.....CCCAATATCCGCTCA..... 520
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
457 AsnLeuThrAlaProAlaSerThrGlyProThrValSerThrHisAspVa 473
521 ..ACGTGACCGACAGCGAGACCGGACACGCTTCCAGCGTTTCCA 569
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
473 lThrSerProThrProAlaGlyThrThrSerGlyAlaSerProValThrPr 490
570 CAATACCGGTAGTATGCTGACGCAAGAGTAGGACGATTCAAACGCG 619
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
490 roSerProSer.....ProArgAspAsnGlyThrGluSerLysAla 503
620 CCACCCGATACAGCCCG...AGCTGACAGATCGGGCAATGCCCGCGAA 666
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
504 ProAspMetThrSerProThrSerAlaValThrThrProThrProAsnAl 520
667 GCTTTCAACGCACTGCAGATATCGTCAAAACATATGCGCGCGGCG 716
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
520 alThrSerProThrProAlaValThrThrProThrProAsnAlaThrSerP 537
717 AGAATTTGTGCGCAGCGAGCGATCCGTGACGAGGTATAGCAAGCTCA 766
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
537 roThrLeuGlyLysThrSerProThrSerAlaValThrThrProThrPro 553
767 ACATTCGCTTATGCAAGCGCTTGCTGCTTCCACCCAAACAAGATG 816
      :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
554 AsnAlaThrSerProThrProAlaValThrThrProThrProAsn..... 568
817 GCGCGCATCAACGATTTGGCAGATATGCGCACAATCA..... 853
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
569 .....AlaThrIleProThrLeuGlyThrSerProThrSerAlaVal 583
854 .....AAGACTATG 862

```

```

583 a1ThrThrProThrProAsnAlaThrSerProThrValGlyGluThrSer 599
863 CCGGACGACGATCCGGGATTT.....GGGACGTCGCAAAACCC 900
600 ProGlnAlaAsnThrThrAsnHisThrLeuGlyGlyThrSerSerThrPr 616
901 AATG.....CCGCACAAAG..... 913
616 ovalValThrSerProProLysAsnAlaThrSerAlaValThrThrGly 633
914 .....GCATAGAACCCGTCAGCAATATCTTACGGCAGCA 949
633 InHisAsnThrThrSerSerThrThrSerSerMetSerThrArgProSer 649
950 TCCCGGTCGCAAGGATGAGCTGTTCGGGAAATAACGCTTGCGGCGC 999
650 SerThrSer.....GluThrLeuSerProSerThrSerAspAsnSe 663
1000 ATCAGCGCAC.....ATCCTGTCAAGCGGTCGACAGATGGCGGATCCG 1043
663 ThrSerHisMetProLeuThrSerAlaHisProThrIlyGlyGlu 680
1044 ATTGGCGAAGGGAATCCGCGCTCA...GGCACAATTTTGGCATGCGG 1090
680 snlThrThrValThrProAlaSerThrSerThrHisValSerThr 696
1091 CATACGCCCAATATCCGTCCTTCATTCGCAATATCCGT..... 1134
697 SerSerProAlaProArgProGlyThrThrSerGlnAlaSerGlyProG 713
1135 ....TCMAACTGTGACGACGCTTACGGCAAGAAACATCCCTCCCTCA 1180
713 YAsnSerSerThrSerThrLysProGlyGluValAsnValThrIlyGly 730
1181 CCGTCCGCGCTCAAAACGGAAGAAATGTGAACCTGGCAACCAACGCCAC 1230
730 hr...ProProLysAsnAlaThrSer..... 737
1231 CCGAAGACCAAAAGTCCGTTTACGCGTAAGGTTTCCGAAATTTGAAA 1280
738 ProGlnAla.....ProSerGlyGlu 745
1281 AGACGTAAATAGATACGAGATTAATACCGCTGTACACAGTAATC 1330
745 s.....ThrAlaValProThrValThrs 753
1331 CTAATAGTAAACCCGCTTTAATCTTAAGGTTCTGTGCGATCGCTCAT 1380
753 eThrGlyGly.....LysAlaAsnSerThrThrGlyGlyIlyShis 766
1381 TCCTGGTCATTAACGTCAGAAATTCATACGCAAAATTCACCAAGCAGG 1430
767 ThrThrGlyHisGlyAlaArgThrSerThrGluProThrThrAspGly 783
1431 TAGAATCAGATATATCCACCTAATAATTAATCTCTCTCAGACGCGTAC 1480
783 GlyAspSerThrThrProArgThrArgThrAsnAlaThrThrIlyLeu 800
1481 CAAAAGACCTATATGATGATTTGGATAAATTTGGTAATGATGAGTACT 1530
800 roProSerThrSerSer.....LysLeuArgProArgThrPr 812
1531 ....AAAGTCCATCAAGAACTAAAGTCAAGAAATTTGAATGGAGTGTCA 1577
813 PheThrSerProProValThrThrAlaGlnAla.....ThrValPr 826
1578 ATTGCTAAACAGAGAGAGCA.....C 1603
826 ovalProThrSerGlnProArgPheSerAsnLeuSerMetLeuVal 843
1604 TTGATGGGCTACT 1617

```

843 euGlnThrProAlaSer 847

```

seq_name: p1r2:S50832
seq_documentation_block:
  atrophin-1 - human
  C:Species: Homo sapiens (man)
  C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 07-May-1999
  C:Accession: S50832
  R:Nagatani, S.; Yanagisawa, H.; Ohsaki, E.; Shitayama, T.; Tadokoro, K.; Inoue, T.;
  Nature Genet. 8, 177-181, 1994
  A:Title: Structure and expression of the gene responsible for the triplet repeat d150
  A:Reference numbers: S50832; MUID:95144175
  A:Accession: S50832
  A:Status: preliminary
  A:Molecule type: mRNA
  A:Residues: 1-1184 <NAG>
  A:Cross-references: EMBL:D31840
  C:Genetics:
  A:Gene: GDB:DRPLA; B37
  A:Cross-references: GDB:270336; OMIM:125370
  A:Map position: 12p13.31-12p13.3112p-12p

```

```

alignment_scores:
  Quality: 131.50      Length: 684
  Ratio: 0.496        Gaps: 39
  Percent Similarity: 38.743      Percent Identity: 22.368

```

```

alignment_block:
  US-09-303-518D-465 x S50832 ..

```

```

Align seg 1/1 to: S50832 from: 1 to: 1184

```

```

65 CACACGCGTCAGATTGGCAACGATCTTTATCCGAGGTTCTCGAC 114
253 HisProPro.....ProThrThrProIleSerValSerSerGly 266
115 CGTCAGCATTTTCGAAACCGACGAGGAAATACCACTTTCGCGACGAGG 164
266 YAlaSerGlyAlaProProThrLysProProThrThrProValGlyGly 283
165 GGAACCTGCGGACGCGCGGTCATATCGGATGGGAAACATCAAGACC 214
283 LysAsnLeuProSerAlaProProProAlaAsnPhe.ProHisValThrPr 299
215 ATCAGTTGGGCAACCTGTCATCCAGCAGG.CGGCAATTAAGAAATAT 263
299 oAsnLeuProProPro.....ProAlaLeuArgProLeuAsnAlaAs 314
264 CGGCTACA.....TTGTCGCGCTTTTCCGATACGCGGACGAGAGTCC 304
314 eAlaSerProProGlyLeuGlyAlaGlnProLeuProGly..... 327
305 ATTCCCGCTTCGACACCAATGCCCTCACATTCGATTCGATGAAGCCGT 354
328 .....HisLeuPro..... 330
355 AGTCCCGTTGACGAGATTCAGCCTTTACCGCATTCATGGGACGATACGA 404
331 .....SerProTyr.....AlaMetGlyGlnGlyMetG 340
405 ACACCATCCGCGCGGCGCTATGACGGGCGACAGGGCGGCGTATCCCG 454
340 IlyGlyLeuProProGlyProGlyLysGlyProThrLeuAlaProSerPro 356
455 .....CTCCCAAGGCGCGAGGATATATACAGCTACGACATAAAGCC 498
357 HisSerLeuProProAlaSerSerSerAlaProAlaProPro...MetAr 372
499 GTTGCCCAAAATATTCGCGCTTACCTGACGACGACGACGACGACGACA 548
372 gpHeProTyrSerSerSerSerSerSerSerAlaAlaAlaSerSerSers 389

```

549 ACGGCTTGCACGCTTTCACAAATACGGTAGTATGCTGACGCAAGAG 598
 389 erSerSerSerSerSerSerLaserPrope..... 399
 599 TAGGCGAGGATTCAACGCCGCCACCGATACGCC...CCGAGCTGAC 645
 400ProLaserGlnAlaLeuProSerTyPr 409
 646 AGATCGGCAATGCCGCCGAGCTTTCACGCACTGCGAGTATCGCA 695
 409 ohSerSerProProPro.....ThSerLeuSerValSerSng 423
 696 AAACATCATCGCGCGCGCGCAAAATTTGCGCGCAGCGCATCCGTCG 745
 423 InProProLysTyThrGlnProSerLeuProSerGlnAlaVal..... 437
 746 AGGATATAGGGAAGGCTCAACATGCTGTATGACAGCGCTTGCGTCTG 795
 438TrpSerG 440
 796 CTTTCACGCAAAACA.....AGATGCGCGCATCAACGATTT 833
 440 nglProProProProProProTyGlyArgLeuAlaAsnSer.... 455
 834 GCGAGATATGCGCGCACTCAACACTATGCCGCGCAGCGCATCCGCGATT 883
 456AsnAlaHisProGlyProPheProProSerThrGly 467
 884 GCGCATGCAAAACCCCAATGCCG.....CACAAAGCATAGAA 921
 468 AlaGlnSerThrAlaHisProProValSerThrHisHisHisHisG 484
 922 GCGCTGCAATATCTTACGCAATCATCCCGCAAAAGGATGAGC 971
 484 InGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 500
 972 TGTTCGGGAAATACGGCTTGGG..... 995
 501 AsnSerGlyProProProProGlyAlaPheProHisProLeuGlnGly 517
 996CGGCATCAGCGCATCT..... 1014
 517 ySerSerHisHisHisHisProTyAlaMetSerProSerLeuGlySer 533
 1014 1014
 534 LeuArgProTyProProGlyProAlaHisLeuProProHisSerG 550
 1015 .GTCAAGCGGTGCGAGATGCGGAGATGCCATTTCCGAAAGGAATCCG 1063
 550 nValSerTySerGlnAlaGlyProAsnGlyProProValSerSerSers 567
 1064 CC...GTCAAGCACAATTTGCCGATGCGCATAC..... 1095
 567 exAsnSerSerSerSerThrSerGlnGlySerTyProCysSerHisPro 583
 1096GCCAAATACCGCTCCCTTACCATTG 1121
 584 SerProSerGlnGlyProGlnGlyAlaProTyProPheProProValP 600
 1122 CCGAATA.....TCCGTTCAACTGCGAGCAGGTACGGCA..... 1159
 600 roThValThrThrSerSerAlaThrLeuSerThrValIleAlaThrVal 616
 1160AAGAAACATCACCTCTCAACCGTGGCG 1188
 617 AlaSerSerProAlaGlyTyTyTySerThrAlaSerProProGlyProPro 632
 1189 CCGTCAACGGAAGAAATGTGAACCTGCAACAAACGCCACCG..... 1233
 633 Pro.....TyGlyLysArgAlaProSerPr 641

1234AAGACCAAGTCCGCTTTGACGGTAAAGGTTT..... 1266
 641 oglyAlaTyLysThrAlaThrPro.....ProGlyTyLysProG 655
 1267CCGATTTTGAAAGAGGTAAGTATGATGACGAGA..... 1302
 655 lSerProProSerPheArgThrGlyThrProGlyTyThrArgGlyThr 671
 1303ATTAAATCCGCTGTACACAGTGA 1328
 672 SerProProAlaGlyProGlyThrPheLysProGlySerProThrValG 688
 1329 TCCTATAGATGAACCGCTTTATCTTAAGT..... 1362
 688 yProGlyProLeuProProAlaGlyProSerGlyLeuProSerLeuPro 705
 1363TCTGCGATGCGCTCATTTGCTGTAACTGCGACA 1401
 705 roProProAlaAlaProAlaSerGlyProProLeuSerAlaThr..... 719
 1402 ATTCAATACGCAAAATTTACCAAGGCAAGGTAGATCATATATGCCACC 1451
 720GlnIleLysGlnIleProAl 726
 1452 TAAATATAC...TCTCTTCAGCACCGCTACCA.....AAGAGACCTA 1492
 726 agLysGlyTyGlyThrProGlySerProValProAlaArgSerProS 743
 1493 AT..... 1494
 743 erProProProLysValValaSerValProSerHisAlaSerGlnSerAla 759
 1495AATGATATTGGATAA..... 1512
 760 ArgPheAsnLysHisLeuAspArgGlyPheAsnSerCysAlaArgSerAs 776
 1513TTGGTAAATGAATGACCTAA...GSTCCAT 1540
 776 PleuTyThrPheValProLeuGlnGlyProSerThrProArgSerGlyPro 793
 1541 CA..... 1542
 793 hTriTPArgArgCysGlyAlaGlyArgAlaAlaArgAlaArgArgLys 809
 1543AGACTAAAG 1553
 810 GlyAlaArgAlaArgAlaGlyAlaGlyLysGlyArgGlyLysG 826
 1554 TCACGAATTTGAATGGAGTTCATTTGTCTAAACAGAGAGA 1596
 826 uArgAspValGlnArgSerValLysLeuAlaGlnGlnGlyArg 840

seq_name: p1r2:T38459

seq_documentation_block:
 hypothetical divergent repeat-containing protein - fission yeast (Schizosaccharomyces
 C:Species: Schizosaccharomyces pombe
 C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
 C:Accession: T38459, T38380
 R:Harris, D.; McDonald, S.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
 submitted to the EMBL Data Library, February 1996
 A:Reference number: 221794
 A:Accession: T38459
 A:Molecule type: DNA
 A:Residues: 1-1748 <HAR>
 A:Cross-references: EMBL:269368; PIDN:CA93290.1; GSPDB:GNO0066; SPDB:SPAC27F1.01c
 R:Klean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.
 submitted to the EMBL Data Library, September 1997
 A:Reference number: 221790
 A:Accession: T38380
 A:Molecule type: DNA
 A:Residues: 1457-1794 <MCL>
 A:Cross-references: EMBL:270691; PIDN:CA94638.1; GSPDB:GNO0066; SPDB:SPAC25G10.09c

A: Experimental source: clone c25610
 C: Genetics:
 A: Gene: SPAC27F1.01c; SPDB: SPAC25610.09c
 A: Map position: 1

alignment_scores:
 Quality: 131.50 Length: 520
 Ratio: 0.632 Gaps: 25
 Percent Similarity: 40.000 Percent Identity: 22.500

alignment_block:
 US-09-303-518D-465 x T38459

Align seg 1/1 to: T38459 from: 1 to: 1794

```

114 CCGTCAGCATTTTCGACCCGACGGGAAATACACCTATTTCGGCAGCAGG 163
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1326 ProSerThrThrSerThrSerPheAsnThrAlaProIleProGlnGlnAl 1342
164 GGGAACTTCCGAGCGCAGCGGTATCGGATTCGGAACATTCAGC 213
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1342 aProLeu.....GluAsnGlnPheSerLysM 1351
214 CATCAGTTGGCAACCTGTCATCCAGCAGG.....CGGCA 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1351 etSerLeuGlnProProValArgProAlaValProThrSerProLysPro 1367
251 TTAAAGAAATATCGGCTACATTCGCGCTTTTCGATCAGGGCAGCA 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1368 GlnIleProAspSerSerAsnValHisAlaProPro.....Pr 1380
301 GTCATTCCTCCCTTCGACACACATGCTT..CACATTCGATTCGTATGA 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1380 oProValGlnProMetAsnAlaMetProSerHisAsnAlaValAsnAla 1397
348 ACCCGGTAGTCCCGTTGACGATTCAGCCTTT..... 379
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1397 rgProSerAlaProGlnArgArgAspSerPheGlySerValSerSerGly 1413
380 .....ACCGCATTCATTCGCGAGCATCGAACACCATCCG..... 415
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1414 SerAsnValSerSerIleGlnAspGlnThrSerThrMetProLeuLysAl 1430
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1430 aSerGlnProThrAsnProGlyAlaProSerAsnHisAlaProGlnVal 1447
450 TCCCG.....CTCCCAAGGCGCGA 469
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1447 alProProAlaProMetHisAlaValAlaProValGlnProLysAlaPro 1463
470 GGGATTTATACAGCTACGACATTAAGCGGTGCCCAAAATATCCGCTC 519
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1464 GlyMetValThrAsnAlaPro.....AlaProSe 1473
520 AACCTGACGACACCGCAGACCGGACAGCGCTTGTGACCGTT.... 565
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1473 rSerAlaProAlaProIleAlaProValSerGlnLeuProProAlaValP 1490
566 .....TCCACAATACCGGTAGTATCGTACGCAAGAG 598
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1490 roAsnValProValProSerMetIleProSerValAlaGlnGln..... 1504
599 TAGCGACGATTCAAACGCGCCACCGATACAGCCCGAGCTGACAGCA 648
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1505 .....ProProSerSerValAlaProAlaThrAlaProSerSerThr... 1518
649 TCGGCAATGCGCGCAGAGCTTTCAACGCGCAGCTGCAGATATGTCAAAA 698
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1519 .....LeuProProSerGlnSerSerPheAlaHisValProSerProA 1533
699 CATCATCGCGCGCGCAGAGAAATTCGCGCGCAGCGATTCGCGCAGG 748

```

```

1533 la..... 1533
749 GTATATAGCAAGGCTCAAAACATGCTGTTATGCACGGCTGGGTCTGCTT 798
1533 ..... 1533
799 TCCACCGAANAACAGATGCGCGCATCAAGATTGGCAGATATGCGCA 848
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1534 .ProPro..... 1535
849 ACTCAAGACTATGCCGACGACCATCCGCGATTCGGCAGATCCAAAC 898
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1536 .....AlaProGlnHisProSerAlaAlaAlaLeuSerAla 1548
899 CCA.....ATCCGCAAGCA..... 916
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1549 ProAlaAspAsnSerMetProHisArgSerSerProTyrAlaProGlnG 1565
917 .....TAGACCGCTCAGCAATATCTTTACGCGACATCCCGCTCA 959
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1565 uproValGlnLysProGlnAlaIleAsnAsnIleAlaProAlaThrAsn 1582
960 AGGATTTGAGCGTTCGGGAAATACGCTTGGG.....CGGCATCA 1003
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1582 emGlyThr.SerGlnSerPheSerProArgMetGlyProValAsnAsn 1598
1004 CGGCACATCTCTCAAGCGGTGCGCAGATGGCGAG.....ATCGCAT 1047
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1598 rGlySer.ProLeuAlaMetAsnAlaIleGlyGlnProSerLeuAlaVal 1614
1048 CCGAAGGGAATCCCGCTGCGACGATTTTCGCGATCGCGCATACCC 1097
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1615 ProAlaValProSerAlaProSerAsnHisPheAsnProPheAlaLys 1631
1098 CAATATCCGCTCCCT.....TACCATTCGCGAAATATCCGTTCA 1138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1631 tGlnProProAlaProSerProLeuGlnProSerGlnHisAspSerAsp 1648
1139 ACTTGAGCAGCGCTTACGCAAGAAACATCCCTCCCAACCGCGCG 1188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1648 snTrpSerGlnHisGlyAspGlnGlnGlnLysAspSerGlnAspAsp 1664
1189 CCGCAACCGCAAGAAATGTGAACCTGGCAACCAACGCCCAACCGCAAG 1238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1665 ArgSerSerLysAspAlaAlaAlaLeuAlaAlaLysLeu..... 1677
1239 CAAAGTCCCTTTCGACGCTAAGGTTTCCGAATTTGAAAAAGACTCA 1288
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1678 .....PheGlyGly..... 1680
1289 AATACGATACGAGATTAATACCGCTGTACCAAGATGATCTATAGAT 1338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1681 .....MetAlaProAlaHisProValSer 1688
1339 GAAACCGCTTTTAATCTTAAGGTTCTGTGCGATCGCTATTCTTGTC 1388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1689 ThrProProValArgProGlnSerAlaAlaProProGlnMetSer.... 1703
1389 TATTAAGTCCAGAAATTAATACGCAAAATTAACCAAGCAGATGATCA 1438
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1704 .....AlaProThrProProProProProMet 1713
1439 GATATATCCGACCTAAATAATTAATCTCTTCAGCA...CCGCTACCAAA 1485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1713 erValProProPro.....ProSerAlaProProMetProAla 1725
1486 GGACCT 1491
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1726 GlyPro 1727
seq_name: pir2:G84693

```

seq_documentation_block:
Probable protein-rich protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: G84693
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentho, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moitel, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Meriman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: G84693
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-891 <STO>
A:Cross-references: GB:AE002093; NID:g3980411; PIDN:AAC95214.1; GSPDB:GN00139
A:Gene: At2g29210
A:Map position: 2

alignment_scores:
Quality: 130.00 Length: 394
Ratio: 0.756 Gaps: 17
Percent Similarity: 43.655 Percent Identity: 25.888

alignment_block:

US-09-303-518D-465 x G84693 ..

Align seg 1/1 to: G84693 from: 1 to: 891

```
144 CCACCTATTCGCGAGCGGGGGAACCTTCGCGAGCGCGCATATTCG 193
||| |||::: |||::: ||| |||::: |||
262 PROAGLYAARGARGLEUSERINSERATGARGSERARGSERARGSe 278
194 GATTGGGAACATACAAAGCCATGATTGGGAACCTGTCATCCAGCAG 243
::: |||
278 rValAARGSERLeuserPro..... 285
244 GCGGCATTAAGGAATATCGGCTACATTCGCGCTTTCGATCAGCG 293
|||::: |||::: ||| |||::: |||
286 .....AARGARGARGyleHISerPro...PheARgSERAr 297
294 GCACGAGTCCATCCCTT...GCACACCATGCTTCATCCGAT 340
::: |||::: |||::: |||::: |||
298 SERARGSERProleAARGHISARGARGProthHISGLYArg.. 313
341 CTGATGAAGCCGATGATCCCGTTGACGATTCAGCCTTACCGCAT 390
::: |||
314 .....AARGINSerProAlaProS 320
391 TGGGACGATAGAACACCATCC.....CGCCGACGCTATGACGG 431
||| |||::: |||::: |||
320 eRARGARGARGSERProSerProAlaARGARG..... 333
432 GCCACAGCGCGGCTATCCCGCTCCCAAGCGCGAGGATATATACA 481
||| ||| ||| ||| |||
334 .....ARGSerProSerProAlaARG..... 342
482 GCTACGACATAAAGCGCTTGCCTCAATATCCCTACCTGACGAC 531
||| ||| ||| ||| |||
343 .....ARGARGSerProSerProAlaARGHISARGS 355
532 AACCG.....CAGCAGCGGACAAAGCGCTTTCGACCGCTTCCA 569
::: |||
355 eRProthProProAlaARGHISARGSERProSerProAlaARG 371
570 CAATACCGTAGTATCGTCGACGAGAGTAGCAGCGGATTCACACGCG 619
::: ||| ||| ||| ||| |||
372 AARGHISARGSERProProProAlaARGARG.....Ar 383
620 CCACCGGATACAGCCCGGAGCTGACAGATCGGCGAATGCGCGACGT 669
||| ||| ||| ||| |||
```

```
383 gSerProSerProPro.....AlaARGARGARGSerp 395
670 TTCAACGGCACTGCAGATATGTCACAAACATCATCGCGCGCAGAGA 719
::: ||| |||::: |||
395 roSerProProAlaARGARGARGSERProSerProLeuTYrARGArg 411
720 AAT.....TTCGCGCGAGCGGATCCGTCGACGGGTA 751
||| ||| ||| ||| |||
412 AsnARGSERProSerProLeuTYrARGARGAsnARGSERProle 428
752 TAACGGAAGCTCAACATTCGCTTATGCACGCGCTTGCTGCTTCC 801
::: |||
428 uAlaYsARGlyARGSERAspSerProGlyARGSerProSerProVala 445
802 ACCGAAAAAGATGGCGGCATCAAGATTTGGCAGATATGGCGCACT 851
||| ||| ||| ||| |||
445 lAARGLeuARGAspProthGlyAlaARGLeuProSerProSerIleGlu 461
852 CAAGACTATGCGCGCAGCAGCATCCGATTTGGCAGTCCAAACCCCA 901
||| ||| ||| ||| |||
462 GlnARGLeu..ProSerProProVala...GlnARGLeuProSerPro 476
902 ATGCCGCAAGCATTAGAACCGCTCAGCAATATCTTTACGCGCATATC 951
||| ||| ||| ||| |||
477 ProProARGARGAlaGlyLeuPro.....Se 485
952 CCGGTCAAGGATTTGAGAGCTTCGCGGGAATATCGGCTTGGCGGCAT 1001
||| ||| ||| ||| |||
485 rProProProAlaGlnARGLeuProSerProProProAlaARGAlaGly 502
1002 CACGCGACATCTTCAAGCGTTCGCGAGATGGCGAGATTCGCGCA 1051
::: ||| ||| ||| ||| |||
502 euProSerProMetARGIleGlySerHISAlaAlaAsnHISLeuGlu 518
1052 AAGGAAATCCGCGTCAGCGACA.....ATTTCGC 1083
::: ||| ||| ||| ||| |||
519 SerProSerProSerLeuserProProGlyARGlySylValleupr 535
1084 GATCGGATATGCGCAATACCGCTCCCTTACCATTCGCGAATATCCG 1133
||| ||| ||| ||| |||
535 oSerProProValaARGARGARGSERLeuThProAspGlnuARGV 552
1134 TTCAAACTTGGAGCAGGCTTACGCAAGAAACATCATCCCTCC..... 1176
||| ||| ||| ||| |||
552 alSerLeu.....SerGlnGlyARGHISrHISrProSerHISle 566
1177 .....TCAACCGTCCCGCTCAACGGAAGATGTGAATCTGGC 1217
::: ||| ||| ||| ||| |||
567 LysGlnAspGlySerMetSerProValaARGIlyARGIlySylSerSerPr 583
1218 AAACAAACGCCACCCGAAAGCAAGATGCCG 1248
::: ||| ||| ||| ||| |||
583 oSerSerARGHISGlnSylAslaARGSerPro 593
```

seq_name: pIrf2:T14180

seq_documentation_block:

exit protein - Mycobacterium smegmatis

C:Species: Mycobacterium smegmatis

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 01-Dec-2000

C:Accession: T14180; T14163

R:Zhu, W.M.; Arceneux, J.E.L.; Beggs, M.L.; Byers, B.R.; Eisenach, K.D.; Lundrigan,

Mol. Microbiol. 29, 629-639, 1998

A:Title: Exochelin genes in Mycobacterium smegmatis: identification of an ABC transpo

A:Reference number: Z17906; MUID:98383832

A:Accession: T14180

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1122 <ZHU>

A:Cross-references: EMBL:AF034152; NID:g3421057; PID:g3421058; PIDN:AAC32046.1

R:Yu, S.; Fiss, E.; Jacobs Jr., M.R.

J. Bacteriol. 180, 4676-4685, 1998

A:Title: Analysis of the exochelin locus in Mycobacterium smegmatis: biosynthesis gen

A:Reference number: Z17898; MUID:98389687
 A:Accession: T14163
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 534-1122 <YUS>
 A:Cross-references: EMBL:AF027770; NID:93560502; PID:93560505; PIDN:AAC82548.1
 C:Genetics:
 A:Gene: exit
 C:Superfamily: Mycobacterium tuberculosis probable ABC transporter RV0194; ATP-binding C

alignment_scores:
 Quality: 130.00 Length: 590
 Ratio: 0.568 Gaps: 31
 Percent Similarity: 38.814 Percent Identity: 23.390

alignment_block:
 US-09-303-518D-465 x T14180 ..

Align seg 1/1 to: T14180 from: 1 to: 1122

```

69 CGCCTCAGATTGGCAACGATTCCTTTATCCGCGAGGTTCTCGACCGTC 118
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
261 ArgAlaGluLeuGlySerArg.....GlyProGlyValGlyGlnProAr 275
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
119 AGCATTTCGAACCCCGCGGGAATACCACTATT...CGCGACGAGGGGG 165
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
275 gArSerArGProGluArGProArSPrProLeuProArGLeuArGlyG 292
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
166 GAACCTGGCGGCGGCGGTCATACGATTCGGAACATACAAAGCA 215
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
292 lYProProArGArG..... 297
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
216 TCAGTTGGCAACCTGTTTCATCCAGCGGCGCATTAAGGAATATCG 265
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
298 .....GlyProProValGlyProGln.....ArgAsnThrAr 308
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
266 G...CTACATTGTCGCTTTTCGATCAGCGGACGACGATCCATTCGCC 312
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
308 gElSerGluYsProValAlArGArGAlArG..... 320
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
313 TTCGAAACCATGCTCACATTCGATTCGTATGAAGCGGTAGTCCGT 362
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
320 ..... 320
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
363 TGAAGGATTCAGCCTTTACCGCATTCATTCGAGCGATACGAACACATC 412
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
321 ...ArgIleArGArGelyThrHisGlu.....ArgArGProAr 332
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
413 CGGCGACGCGCTATGACGGGCGACAGGGCG...CGGCTATCCGCTCC 459
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
332 gCysArGArGThrHisArGSerAlaLeuArGIlleSerAlaSerArGlyG 349
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
460 AAAGGCGGAGGGAATATATACAGCTACGACATAAAGGATGCCCAAA 509
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
349 lArThrHisGlyAla.....GlyArGThrProAla 358
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
510 TATCCGCTCAACCT...GACCGACACCGCGACCGGACGACCGCTTG 556
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
359 ProArGProArSPrProAlaGluArGAlaPro.....ArgThrProAlaAr 373
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
557 TCGACCGTTTCGAATACCGGTAGTAGCTGACGCAAGAGT..... 599
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
373 gArGThrProProGlnLeuValAlGlnArGSpAlaGlyAsnGlnSerG 390
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
600 .....AGGGACGAGATTCAAGCGCGACCGCATACAGCCCGGACT 641
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
390 lAcHisArGAsnArGSerAlaAlaArGThrGlyLeuArG...ArgArG 405
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
642 GGACAGATCGGCAATGCGCGGAGGCTTTCAGCGCACTGACGATATCG 691
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
406 ArgThrThrGlyValAlArGArGArGArGArGArGProAlaProArGAr 422
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

692 TCAAAAACATCATCGCGCGGACAGAGAATTCGCGGACGCA.... 737
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
422 gSerArGProCysArGArGlyAlaArGcysGluProValAlArGArGThr 439
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
738 .....TGGCGTCAGGGGTATAGCGCAAGGCTC 764
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
439 hrProThrLeuArSPrProGlyProCysAlaAlaGly..... 450
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
765 AAACATTGCTGTATGACAGGCTTGCGTTCGATCCAGCAAAACAGA 814
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
450 ..... 450
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
815 TGCGCGCATACACGATTTGGCAGATATGGCGCACTCAAGACTATGCC 864
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
451 .....ArgSerGlyArGProGlyAla.....Ala 459
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
865 GCACGACCATCCCGCGATTTGGCGAGTCCCAAAACCCCAATCCGCGACAG 914
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
459 rGSerArSPrHis.....ArgArGArG 465
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
915 CATGAAGCCGT.....CAGCAATATCT 937
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
466 HisArGHisArGAlaGlyHisArGArGArGHisThrProAlaAlaHisAl 482
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
938 TTACGCGAGTCAT.....CCCGCTCAAGGAGATTGAGCTGTCCG 978
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
482 aArGlyThrHisAsnGlyGlyProHisGlnGln.....SerG 495
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
979 GGAATATACGGCTTGGGCGCA.....TCACGCGCATCTGTCAAGC 1021
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
495 lYProAlaAlArGArGlyArGProGlyAlaAlaAlArGHisArGAlaSer 511
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1022 GGTGCGAGATGGGCGAGATGCGATTCGCGAAGGAAATCCGCGTACG 1071
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
512 GlyGlnArG..... 514
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1072 GACAATTTTCCGATGCGCGCATACGCAATATAC.....CGTCCCTTA 1115
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
515 .....ArgHisThrProArGSerGlyArGArGProLeuAla 526
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1116 CCATTCCGGAATATCCGTTCAAACTTGAGCGAGCGATTACGCAAGAA 1165
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
526 rGLeuProGlySerGlyAsp.....PrometIlleSerArSPrGluPro 539
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1166 ACATCACCTCTCTCAACCGTCCGCGCTCAACGGAAGAAATGTGAACATG 1215
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
540 ValThrProProValLeuProValAlaThrAlaArGArGSerGlyAlAr 556
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1216 G.....CAAAACAGCGCACCGCAAGACCAAGAGTCCGTTGACG 1256
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
556 rLeuIlleAlaAsPrLeuArGlyArGArGThrAlaLeuAlaAlaValVal 573
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1257 TAAAGGTTTCCGAATTTGAAAAAGACGTAAATACGATACAGAAATTA 1306
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
573 hrValGlyLeuAlaAla.....AlaAlaAlaSerVal 583
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1307 ATACCGCTGTACCAAGATGATCTATAGATGACCCGCTTAAATCT 1356
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
584 ValProIleTy.....Leu 589
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1357 AAAGTTCTGTGGATGGCTCAT..... 1381
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
589 uGlyMetLeuValAsPrArGValGlnHisGlyAsPrGlyThrArGlyLeu 606
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1382 .....CTTGGT 1387
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
606 alAlaLeuGlyThrValIlleAlaLeuAlaAlaValAlaGlyGlyLeuGly 622
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1388 CTATTACTGCCAATTCATACGCAAAATTTACCAAGCAAGGTAGAATC 1437
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
623 ThrGlyLeuSerThrTyrlleuThrSerArGLeuGlyGlyGlnMetLeuAl 639
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1438 AGATATATTC.....CACTTAAATAATTATCTC 1463
   ||| :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```



```

639 aaSPLeuArgLuarValLeuGluValAlaLeuAsnLeuProAlaThrL 656
1464 TCCTTCAGCAGCGCTACCAAAAGACCTAATATGATATTGATGAAT 1513
656 euleGluGluSerGlyArgGlyAspLeuSerArgValAlaGlyProasp 672
1514 TTGGTAATGAATGACTAAAGTTCATCAAGACTAAAGTCAAGAAATT 1563
673 ValAlaValAlaArgThrValAlaGlnValLeuProVal...IleLe 688
1564 GAATGGATTTCAATTGCTTAAACAGAGAGACCACTGGATGGC 1613
688 uasnGlyPhePhe.....LeuGlyValValThrL 698
1614 TAGTAGGATGTAGACATT 1633
698 euValGlyMetThrThrLeu 704

seq_name: p1r2:T51023

seq_documentation_block:
hypotheical protein B7F21.40 [imported] - Neurospora crassa
C.Species: Neurospora crassa
C.Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C.Accession: T51023
R.Schulte, U.; Alyn, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, July 2000
A.Reference number: Z25286
A.Accession: T51023
A>Status: preliminary
A.Molecule type: DNA
A.Residues: 1-2649 <SCH>
A.Cross-references: EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.40
A.Experimental source: BAC clone B7F21; strain OR74A
C.Genetics:
A.Gene: NCSP:B7F21.40
A.Map position: 6
A.Introns: 1619/3; 2584/1

alignment_scores:
Quality: 130.00 Length: 593
Ratio: 0.551 Gaps: 33
Percent Similarity: 39.798 Percent Identity: 21.585

alignment_block:
US-09-303-518D-465 x T51023 ..

Align seg 1/1 to: T51023 from: 1 to: 2649

236 TCACAGCAGCGCCATTAA.....AAGAAATATCGGCTAC 270
1961 SerThnGlnArgProMetProThrSerMetGlnGlnIlySerAlaAs 1977
271 ATTGTCCGCTTTTCGATCAGCGGACGAAGTCATCCCTTGGACAA 320
1977 nhtrSerAla.....AlaGlnProValProProGlnA 1989
321 CCATGCTTCATTCGATTCGTATGAGCCGTAGTCCCGTTGACGAT 370
1989 lAProProAlaIlePro.....ProGlnProLysLysValSer 2001
371 TCAGGCTTACC..... 382
2002 SerIlePheSerIleLeuAsnAspAsnProProAlaProAlaProAl 2018
383 .....GCATCCATT.....GGGAGGATGAGA 404
2018 aProLysArgValAlaAspValAlaIleSerMetProArgAlaIleSerThn 2035
405 ACACCATCCCGCCGACGGCTATGAGCGGCAC..... 436

```

```

2035 eThrProProGlnGlnMetSerAlaArgProProGlnProProPro 2051
437 ...AGCGGGGGGCTATCCCGTCCCAAGCGCGAG.....G 471
2052 ProThrThrAlaValSerAlaSerGlnArgAspThrGlnAlaLeuG 2068
472 GATATATACAGTACGACATTAAGCGCTGCCCAAAATATCGCT... 518
2068 lYThrSerThrAlaArgAsn.....ProProSer 2077
519 .....CAACTGACGACACACCGACCGACCGACAGCGC 553
2078 AlaAlaGlnAlaAlaMetProSerLeuLysProThrIleThrGlnSer.. 2093
554 TTGTGACCGCTTCCACATACCGGTATGCTGACGCAAGAGTACGC 603
2094 .....ProGlnProProHisMetAsnValAla..... 2102
604 GACGATTCAAACCGCGCCGATACAGCCCGACGCTGACAGATCGGG 653
2103 ..ArgSerSerMetGlySerGlyMetGlnProGlnValSerAlaGly 2118
654 CAATGCGCGCAAGCTTTCAACGCGACTGCAGATATCGTCAAAACATCA 703
2119 AspAsnArgGluPheThrSerArgHis...GlnThr...GlnProGlnH 2133
704 TCGGCGCGGAGAGAAATTTGCGCGGACGCGATCCCGTACAGGATATA 753
2133 SGIProGlyAlaSerAsn..... 2139
754 AGCGAAGGCTCAACATTTGTTATGACGCGCTTGGGTCTGTTTCAC 803
2140 ..SerProValProHisGlnValHisThrIlySerGlnSerAlaGlnHis 2155
804 CGAAACAGAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCACTCA 853
2156 ProGlnGln.....HisLeuGln.GlnProGlnMetAlaThr.Pro 2168
854 AAGACTATGCGCGCAGACGCGCATCCGCGATGGGCGAG.....TCCAAAAC 897
2169 SerGlnGlnGlnIlyThrProThrAlaThrSerGlnAlaLeuAlaIle 2185
898 CCCAATGCCGCGCACAAGCATAGAAGCCGTCAGCAATATCTTTCAGCGAGT 947
2185 rProThrProGlnIlyThrAlaAlaHisProSerSerIleSerGlyArgArg 2202
948 CATCCCGCTCAAGAGGATTTGAGCTGTTCCGGGCAAAATACGGCTGGCG 997
2202 luaAlaGlnSerAlaAlaArgGlnIlyThrProSerSerThrGlnGlnAla 2218
998 GCATCAGCGCACATCTGTCAAGCGGTGCGAGA..... 1030
2219 AlaValSerGlyLeuGlnGlnArgGlnGlnGlnGlnGlnGlnGln 2235
1030 ..... 1030
2235 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 2252
1031 .....TGGCGGATTCGCAT.....TGGCGAAG 1054
2252 lGlnGlnGlnIlyThrProProSerHisProThrProSerLysSerSerGln 2268
1055 GGAATCCCGCGCTGAGGACAAATTTTCCAGTGCGGCGCATGCCAATAC 1104
2269 AlaSerValProSerGlnIlyThrAlaThrAlaAlaGlnHisGlyProAsn.. 2284
1105 CCGTCCCGCTTACATTCGCCGAATATCCGTTCAAACTTGAGACAGGCTTA 1154
2285 .....ValG 2286
1155 CGGCAAGAAACATCACTCTCAACCGTGCAGCGCTCAACAGGAAGA 1204
2286 lnaIlyAs.....ProProGlnMetGlySerAlaMetSerGlnGln 2299

```

```

1205 ATGTAAATGCGCAACAAACGCCACCCGAAAGCCAAAGTCCGTTGAC 1254
      ::::::::::::::::::::
2300 GlnHisSerTrpGlnAlaThrProThrGlnGlnPro..... 2311
1255 GGTAAAGGTTCCGAATTTTGAAGAACGCTAAATATGATACAGAGAT 1304
      ::::::::::::::::::::
2312 .....HISA 2313
1305 TAATACCGCTGTACCAAGTGAATCTATAGATGACCCGCTTTT... 1350
      :::::
2313 laLeuGlyLeu..ArgGlnProAlaProAlaGlyGlnAlaValaPheSerA 2329
1351 .....AATCTAAAGGTTCTGTGCGATCG...GCTATTCT 1383
      :::::
2329 laHisGlyAlaGlnSerProThrGlySerValaValaSerHisGlnHis 2345
1384 TGGTCAATACCTGCCGAATTCATACGCAAAATATACCA..... 1422
      :::::
2346 ArgSerLeuAspGlyArgSerGlnPheProPromeProAspProAlaGAs 2362
1423 .....AGCAAGCTAGA..... 1434
      :::::
2362 parGlnAsnLeuArgArgGlyGlnProValaProProGlnGlnProT 2379
1435 ..ATCAGATATATC.....CCACCTAAATATCTCTCTCTCAGCA 1473
      :::::
2379 yValaArgTyValaAsnThrProGlyProGlyHisGlyGlyProGlyGly 2395
1474 CCGCTACCAAAAGACCTAATATGATATTTGGAT..... 1509
2396 ProProGlyGlnValaProGlyArgGlyGlnAlaProAlaAspLeuArgme 2412
1510 .....AAATTTGGTAAATGAATGACTAAAGTCCATCAGACATA 1549
      :::::
2412 tGlnGlnMetSerGlnAlaArgSerTyrThrProGlyPro.....V 2426
1550 AAGCTCAAGAAATTTGAATG..... 1569
      :::::
2426 aGlyValaGlyPheGlnGlyMetGlyProProSerSerSerLeuGly 2442
1570 .....GATGTCATATGCTTAA 1586
2443 TyrProGlnGlnIleArgAspAlaGlnIleArgAspAlaGlnLeuArgG 2459
1587 AACAGAGAGAGACCACTTGA 1608
2459 uMetGlySerArgAspLeuGly 2466

```

seq_name: p1r2:T12970

seq_documentation_block:
 hypothetical protein T6H20.190 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 22-Oct-1999
 C:Accession: T12970
 R:ChoiSue, N.; Robert, C.; Brottier, P.; Wincker, P.; Catolico, L.; Artiguenave, F.; Sa
 submitted to the Protein Sequence Database, July 1999
 A:Reference number: 217586
 A:Accession: T12970
 A:Molecule type: DNA
 A:Residues: 1-510 <CHO>
 A:Cross-references: EMBL:AL096659; GSPDB:GN00061; ATSP:T6H20.190
 A:Experimental source: cultivar Columbia; BAC clone T6H20
 C:Genetics:
 A:Gene: ATSP:T6H20.190
 A:Map position: 3
 A:introns: 139/3; 295/3; 318/3; 338/2

alignment_scores:
 Quality: 129.50 Length: 544
 Ratio: 0.514 Gaps: 26

Percent Similarity: 46.324 Percent Identity: 20.772
 alignment_block:
 US-09-303-518D-465 x T12970 ..
 Align seg 1/1 to: T12970 from: 1 to: 510

```

242 AGCGCGCATTAAGGAATATGCGTACATTTGCGCTTTCCGATCAC 291
      :::::
23 ArgysProArgLeuThrValITrpAlaLysGlnThrAlaPheGlnLeuG 39
292 GGGCAGCAAGTCCATCCCGCTTGACA..... 319
39 LysThrLysGlyAspAspSerGlnGlyLysGlnLysGlyLysAsn 56
      :::::
320 .....ACCATGCTCAGATTCGATTCGATGAG 349
      :::::
56 ropheGlnPheAspPheGlyLysLeuProAspMetLysSerLeu... 71
350 CCGGTATC.....CGGTGAGGATTCAGCCTTAAACCGCATTCATGG 393
      :::::
72 ProValaThrAsnProSerThrGlyLeuValaPheGlyAsnAsnArgly 88
394 GACGATACGACACCATCCCGCCGACGCTATAGCGGCGCACAGGCG 443
      :::::
88 sLysAspProGlyThrLeuPheValaGlyAlaThrGly..GlnAlaG 104
444 C.....GCTATCCCGCTCCAAAG 463
104 yIleArgIleAlaGlnThrLeuLeuGlnArgGlyPheSerValaArgAlaG 121
464 GCGGAGGATATTA..TACAGTACGACATAAAGCGCTGCCCAAAAT 510
121 lValaProAspLeuGlyAlaAlaGlnAspLeuAlaArgValaAlaAlaThr 137
511 ATCCGCTCAACTGACCGACCAACCGCAGCACCGCAGCAACGCTTGCA 560
      :::::
138 TyrIleLeuSerAsnAspGlyValaLysArgLeuAsnAlaValaGlnse 154
561 CCGTTCCACAAATACCGGTATGCTGACGCAAGAGTATGGCAGGAT 610
      :::::
154 rProPheGlnAspAlaGlnSer...IleAlaLysAlaIleGlyAsn... 168
611 TCANAAGCGGCGGATACAGCCGCGAGCTGAGAGATGGGCAATGCC 660
      :::::
169 .....AlaThrLysValaValaValaThrValaGlyAlaThrGlnAsnGly 182
661 GCCGAGCTTTCAACGCGCAGCTGACAGATATGTCAAAACATCATCGCGCG 710
      :::::
183 ProAspAlaGlnValaSerThrSerAspAlaLeu.....LeuValaG 197
711 GGCAGAGAAATTTGCGCGCAGCGCATGCCGTG.....CAGGTA 751
      :::::
197 naIleAlaGlnLeuAlaGlyAlaSerHisValaIleAlaIleValaTyrAspL 214
752 TAAAGGAAGGCTCAACATGCTGTTATGACGCGCTG..... 789
214 hrIleSerGlySerThrTyrAsnValaLeuAspGlyIleThrSerPhe 230
790 GGTCTGCTTCCACCGAAGAACAGATGGCGCGCATCAGCATTTGGCAGA 839
      :::::
231 GlyAsnLeuPheAlaLysSerGlnProLeuThrIleSerAspLeuIleG 247
840 TATGGCGCACTCAAAACATATGCCGACAGACCATCCGGATTTGGCGAG 889
      :::::
247 ulysValaIleGlnThrAspValaIaTyrThrLeuIleLysThrSerLeu 264
      :::::
890 TCCAA.....AACCCCAATGCCGACCAAGCATAGAACCGTCAGCAAT 933
      :::::
264 hrGluAspPheSerProGlnLysAla..... 272
934 ATCTTACGCGCATTCGCCGTCGTCAAAGGATTCGAGCTTTGGGGAA 983
      :::::

```

```

273 ...TyrAsnValValValSerAlaGluGlySerAsnSer..... 284
984 ATACGGCTTGGGCGCATCAGCGACATCTCTCAAGCGGTGCGAGATGG 1033
285 ...GlySerGlySerSerSerGluAlaTyrIlys..... 295
1034 GCGAGATCGCATTCGCCAAGGAAAA.....TCCGCCGTGCGACAAAT 1077
296 .....ValProLysLeuLysIleAlaSerLeuValAlaIleAspIle 308
1078 TTTCGCCGATCGCGCATACGCCCAATAACCCCTCCCTACCATTCGCCGAAA 1127
309 PheAlaAsnThrAlaValAlaGlu..... 316
1128 TATCGGTTCAAACTTGAGCAGCGCTTACGGCAAGAAAACATCACTCTCT 1177
317 .....AsnLysValValGluValS 323
1178 CAACCGTGCCG..... 1188
323 eThrAspProSerAlaProSerArgProValAspGluLeuPheSerVal 339
1189 ...CCGTCAACGGAAGAAATGTGAACCTGCGCAAC..... 1221
340 IleProGluAspGlyArgArgValTyrAlaAspAlaIleAlaArgG1 356
1222 AAACGCCACCGCAGACCAAGCAAGTCCGTTGAC..... 1254
356 uArgAlaGluGluGluAlaLysValAlaAlaAspLysAlaArgGluAla 373
1255 .....GGTAAAGGTTTCCGAAT.....TTTGA AAAA 1281
373 laGluAlaLysGluPheGluLysGluMetGluLysLeuSerGluLys 389
1282 GACGTAATAATACGATACGAGAATTAATACCGCTGACCAAGTGAATCC 1331
390 GluAlaGluAlaAlaSerLeuAlaGluAspAlaGluGluLysAlaAspAl 406
1332 TTTAGAT.....GAACCGCTCTTAATCTTAAGTCTCTGCGGAT 1372
406 aValGlyValThrValAspGlyLeuPheAsnLysAlaLysAspIleSerS 423
1373 CGGCTCATTCCTTG..... 1386
423 eArgLysLeuSerTrpAsnLysLeuGlySerGlnPheAlaThrAlaIleGln 439
1387 .....TCTATACTGCCAGAAATTCATACCAAAATTTACCAAGCGAAG 1430
440 AsnAlaSerGluThrProLysValGluValAlaThrValAlaArgGlyGlnAl 456
1431 TAGAATCGATATATCCACCTAAATAATTACTCTCTCAGCAGCGCTAC 1480
456 aLysAlaArgAsnLeuProPheLys.....LysAlaValValL 469
1481 CAAAGACCTAATATATGATATTTGATTAATTTGTAATGAATGAGACT 1530
469 ysglnArgProSerSerProPheAlaSerLysProLysGluGluArgPro 485
1531 AAAGTCCATCAAGACTAAAGTCAAGCAATTTAATGGAGTGTCAATT 1580
486 LysLysProGluLysGluValAlaArgLysValAlaPhe.GlyGlyLeuPhe.... 500
1581 GTCTAAACAGAGAGAGAGCAACTTGGAT 1609
501 .....LysGlnGluThrIleTyrIleAsp 508
seq_name: pir2:s59310

```

seq_documentation block:
probable membrane protein YMR317w - yeast (Saccharomyces cerevisiae)
N/Alternate names: hypothetical protein YMR924.09
C/Species: Saccharomyces cerevisiae
C/Date: 29-Nov-1995 #sequence_revision 23-Feb-1996 #text_change 05-Nov-1999

C/Accession: S59310
R/Churcher, C.M.
Submitted to the EMBL Data Library, September 1995
A/Reference number: S59302
A/Accession: S59310
A/Molecule type: DNA
A/Residues: 1-1104 <CHU>
A/Cross-references: EMBL:Z54141; GSPDB:GN00013; MIPS:YMR317w
A/Experimental source: strain Ab972
C/Genetics:
A/Gene: MIPS:YMR317w
A/Map position: 13R

alignment_scores:
Quality: 129.50 Length: 430
Ratio: 0.589 Gaps: 17
Percent Similarity: 51.163 Percent Identity: 22.558

alignment_block:
US-09-303-518D-465 x S59310 ..

Align seq 1/1 to: S59310 from: 1 to: 1104

```

86 ACGATTTCTTTATCCGGCAGGTTCTGACCGTCAAGATTTCGAACCCGAC 135
111 : : : : : : : : : : : : : : : : : : : : : : : : : : :
180 ThrAlaValThrSerSerThrPheThrThr.....LeuThrAspValSe 194
136 GGGAAATATCACCACTATTCGCGACAGGAGGGAACCTTGCCGAGCGGAC 185
194 rSerSerProLysIleSerSerSerGlySerAlaValThrSerValGlyT 211
186 TCATATCGGATTTGGGAAACATACAAAGCCATCAGTTGGGCAACCTGTGA 235
211 hThrSerAsp.....AlaSerLysGluValPheSer 221
236 TCACGACGGGGGCCATTAAAGAA...ATATCGGCTACATTCGCCCTT 282
222 SerSerThrSerAspValSerSerLeuLeuSerSerThrSerSerProAl 238
283 TCCGATACGCGGACGAGACGATTCATCCCTTCGACACACGCTCTCACA 332
238 aSerSerThrIleSerGluThrLeuProPheSerSerThrIle.....L 253
333 TTCGATTTGTGAAAGCCGATGTCCTGACGATTCAGCCTTAC 382
253 eSerIleThrSerSerProValSerSerGluAlaProSerAlaThrSer 269
383 GCATTCATTTGGAGCATACGATACACCATC..... 412
270 SerSerValSerSerGluAlaSerSerSerThrSerSerValSerSe 286
413 ...CCGCGACGCTATGACGGGCCACAGGGCGGCGGTATCCGCTCC 458
286 rglnAlaProLeuAlaThrSerSerValValSerSerGluAlaProSerS 303
459 CAAAGCGCGAGGATATATACAGTACGACATTAAGCAAGCGTTCGCCAAA 508
303 eThrSerSerValValSerSerGluAlaProSerSerThrSerSerSer 319
509 ATATCGCGCTCAACCTGACCGACACGACGACCGGACGAGCGTGTTC 558
320 ValSerSerGluIleSerSerThrThrSerSerSerValSerSerGluAl 336
559 GACCGTTTCACAAATACCGTAGTATGTCGAGCAAGAGAGTGGGCGAGG 608
336 aProLeuAlaThrSerSerValValSerSerGluAlaProSerSerThrS 353
609 ATTCAAC.....GGCCACCGCATACA..... 631
353 eSerSerValSerSerGluAlaProSerSerThrSerSerSerValSer 369
632 .....GCCCGAGCTGACACATCGGCGCAATGCCGCGAAGCTTCAAC 675

```



```

|||||..... |||||..... |||||
498 rThrProGluSerLysAlaGluValPheGlyGluArgGlyLysTrpAlaA 515
999 CATCAGCGCATCTCTGTCAGCGGTCGACGATGGCGCGATCGCATTCG 1048
515 laaSPrhoHsgIntYrAlaGluGlyProArgTrp...ArgSerValIle 530
1049 CGAAAGGAAATCCG.....CCGTACGACCAAT 1078
531 LeuLysProLysGlnThrValPhePheProSerGlyThrVal 544

```

seq_name: p1r2:S24169

seq_documentation_block:

mucin - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 10-Nov-1995

C:Accession: S24169

R:Huan, L.J.; Xu, G.; Forstner, G.; Forstner, J.

Biochim. Biophys. Acta 1132, 79-82, 1992

A:Title: A serine, threonine and proline-rich region near the carboxyl-terminus of a rat

A:Reference number: S24169; MUID:92379096

A:Accession: S24169

A:Molecule type: mRNA

A:Residues: 1-292 <HDA>

alignment_scores:

Quality: 128.50 Length: 342

Ratio: 0.959 Gaps: 14

Percent Similarity: 39.181 Percent Identity: 22.807

alignment_block:

US-09-303-518D-465 x S24169 ..

Align seg 1/1 to: S24169 from: 1 to: 292

```

224 GCAACCTGTTTCATCGACGAGCGGCATTAAAGAAATATCGGTACATT 273
5 SerThrProThrSerGlnThrProThrProThrSerThrThrThrG1 21
274 GTCCGCTTTTCG.....ATCAGGGGACGAATCCATCCCTT 314
21 uileSerThrProThrProHisValThrThrThrThrIleIleGlySerS 38
315 CGACACCAATGCTCATATTCGATTCGTATGTAAGCGGTAGTCCGTTG 364
38 erThrSerThrProThrThrThrThrIleThrGluThrProSerProSer 54
365 ACGGATTCAGCTTACCGCATTCATTGGACGATAGCAACATCC 414
55 ThrIleSerGluThrAlaSerSerThrSerProThrThrGluSerThrSe 71
415 GCGGAGCGGTATGACGCGGCACAGCGCGGCTATCCCGCTCCCAAGG 464
71 rSerThrThrThrThr.....SerGlnThrProThrPro... 82
465 CGGAGGATATATACAGCTAGACATATAAAGCGTTGCCCAAAATATCC 514
83 .....ThrProThrThrIleThrGluThrPro...ThrSer 93
515 GCTCAACTGACCGCAACCGACGACCGGACGCGGTTCGACCGT 564
94 ThrSerThrValProThrThrThrGlySerThrSerSerLysProThrG1 110
565 TTCACAAATACCGGTAGTATGCTGACGCAAGAGTAGGACGAGATCAA 614
110 ySerSerThrProIle..... 115
615 ACGCGCACCCGATACAGCCCGAGCTGACAGATCGGCAATGCGCGCG 664
116 .....ThrSerProSer.....ThrProSerProPro 124

```

```

665 AAGCTTCAACGCGCATGTCAGATATCGTCAAAAACATCATCGCGGCA 714
125 ThrGluSerThrThrIleSerSerThrProValThrThrThrAlaThrSe 141
715 GGAGAAATGTGCGCGCAGCGGATCCGTCAGGGATATAAGCAAGGCTC 764
141 rSerThrThrSerSer..... 146
765 AAACATTGCTGTTATGACAGCGCTTGGCTGCTTTCACCGCAACAGCA 814
147 .....ProGlyThr... 149
815 TGGCGCGCATCAACGATTTGGCAGATATGCGCAACTCAAAAGCATATGC 864
150 .....ThrSerPr 152
865 GCACAGCATCCCGCATTTGGGCGAGTCCCAAAACCCCAATGCGCACAGG 914
152 oPheAlaThrSerSerVal...SerSerThrProProSerPro..... 165
915 CATAGAAGCGGTGACCAATATCTTTACGCGCATGCTCCGTCGCAAGGA 964
166 .....ProSerSerThrSer.....GlyProThrThrSerSerGly 177
965 TTGAGCTGTTTCGGGAAATACGCGCTTGGCGGCGCATCAGCGACATCT 1014
178 MetProThrSerSerLysThrThrThr..... 186
1015 GTCAAGCGGTGCGCAGATGGCGAGATGTCATGCGCAAAAGGAATCCG 1064
187 .....GlyProT 189
1065 CGTCAGGACAAATTTGCGATGCGGCGCATAGCCCAATACCCGTCCTT 1114
189 hrSerProThrThrArgPro.....ProSerThrSerThrPro 201
1115 ACATTTCCCAATATCCGTTCAAACTTGACAGACGCTTACGCAAGAA 1164
202 ThrSerProThrValProThrSerThrThrGluAlaIleThrGlnThrAr 218
1165 AACATCACTCTTCACCGCGTCCGCGCTCAAGAGGAAGATGAACT 1214
218 gIeuSerSerThrThrPro...ThrMetGluThrThrArgThrSerSer 234
1215 GCAACAAACGCCCGCAGAACCA 1240
234 rpglyThrValSerSerAlaSerPro 242

```

seq_name: p1r1:VGBEX1

seq_documentation_block:

glycoprotein X precursor - equine herpesvirus 1 (strain Ab4p)

C:Species: equine herpesvirus 1

A:Note: host Equus caballus (domestic horse)

C:date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jul-1999

C:Accession: H36802

R:Teiford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.

submitted to GenBank, March 1992

A:description: The DNA sequence of equine herpesvirus-1.

A:Reference number: A36805

A:Accession: H36802

A:Molecule type: DNA

A:Residues: 1-797 <TEL>

A:Cross-references: GB:M86664; NID:930791; PIDN:AA02506.1; PID:9330862

R:Teiford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.

Virology 189, 304-316, 1992

A:Title: The DNA sequence of equine herpesvirus-1.

A:Reference number: A41831; MUID:9229556

A:Contents: annotation; possible protein-coding frames

A:Note: neither amino acid nor nucleotide sequence is given

C:Genetics:

A:Gene: 71

C:Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein h

C;Keywords: glycoprotein; transmembrane protein
 F:1-22/Domain: signal sequence #status predicted <Sig>
 F:23-797/Product: glycoprotein X #status predicted <AMT>
 F:23-465/Region: serine/threonine-rich
 F:489-797/Domain: equine herpesvirus 1 glycoprotein homology <EHG>
 F:766-790/Domain: transmembrane #status predicted <TMN>
 F:790/Binding site: carbonylrate (Asn) (covalent) #status predicted

alignment_scores:

Quality: 128.50 Length: 565
 Ratio: 0.547 Gaps: 18
 Percent Similarity: 41.593 Percent Identity: 20.177

alignment_block:

US-09-303-518D-465 x VGBEX1 ..

Align seg 1/1 to: VGBEX1 from: 1 to: 797

```

110 TCAGACGTCAGATTTGCAACCCGAGGGAATAACCACTATTGCGACG 159
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 SerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 38
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 AGGGGGGAACTTGCAGCGCAGCGATATGATTCGGAACAATACA 209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 rGly.....:|||||:|||||:|||||:|||||:|||||:|||||:
210 AAGCCATCATTTGGCAGACCTGTTATTCAGAGGGCCATTMAAGAA 259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40 InsertThrSerSerGlyThrThrThrThrThrThrThrThrThrThr 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
260 ATATCGCATATTGTCGGCTTTCCGATCAGCGGACGAGCAAGTCCTC 309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 ProProThrThrThrThrThrThrThrThrThrThrThrThrThrThr 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 CCCTTGACAA...ACCATGCTTCACATTCGATTGTGATGAAGCCGTA 356
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
73 oSerSerThrThrThrThrThrThrThrThrThrThrThrThrThrThr 90
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
357 TCCCGTTGACGATTCAGCCTTTACCGCA.....TCCATTGGGAGC 397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90 laProSerThrThrThrThrThrThrThrThrThrThrThrThrThrThr 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
398 GATACGACACCAATCCCGCGCAGGCTATGAGGGCCAGAGCGCGCGC 447
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 GluThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
448 TATCCCGCTCCCAAGGCGGAGGATATATACAGTACAGATMAAGG 497
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123 rThrAlaAlaProThrThrThrThrThrThrThrThrThrThrThrThr 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
498 CG.....:|||||:|||||:|||||:|||||:|||||:|||||:
140 laSerThrSerAlaGluThrThrThrThrThrThrThrThrThrThrThr 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
527 CCGACACCGCGAGCAGCGAGCAAGCGTTGTGACCGTTCCACAAATAC 576
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
157 ProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
577 GGTATGATGCTGACGCAAGAGTAGGCGATTCGCAAGCGCCACCGC 626
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
173 rVal.....:|||||:|||||:|||||:|||||:|||||:|||||:
627 ATACAGCCCGAGCTGACAGATCGGCAATGCCCGCAAGCTTTCACG 676
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 sPThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
677 GCACGACGATATGTCMAAAACATCATCGGCGCGGCGAGGAATATTC 726
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
727 GCGCAGGCGATGCGCTGACAGGTATTAAGCAAGGCTCAACATGTCGT 776
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

216 rAlaAlaThrThrThrThrThrThrThrThrThrThrThrThrThrThr 233
777 TATGACGCGTTGGGTGCTGCTTCCACGAAACAAAGGCGGCCATCA 826
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
233 hrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 247
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
827 ACGATTGGCAGATATGCGCACTCAAGACTATGCGCAGCAGCA.. 874
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
248 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
874 .....:|||||:|||||:|||||:|||||:|||||:|||||:
264 rSerGlySerThrSerThrThrGlyAlaSerThrThrProSerAla 281
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
875 .....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
281 eThrAlaThrSerAlaThrProThrSerThrSerThrSerAlaAla 297
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
887 CAGTCCAAACCCCAATGCCGCAAGCA..... 916
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 ThrThrSerThrProThrProThrSerAlaAlaThrSerAlaGluSer 314
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
917 .....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 rThrGluAlaProThrSerThrProThrThrAspThrThrThrProSer 331
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
930 CAATATCTTTAGCGCAGATCCCGTCMAAGGATGAGACTGTTCGG 979
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
331 lAlaThrThrAlaThrThrSerPro..... 339
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
980 GAAATACGGCTTGGCGCATCAGCGCACATCTGTCAGCGGTGCGAG 1029
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
340 GluSerThrThrValSerAlaSerThrThrSerAlaThrThrThrAla 356
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1030 ATGGCGAGATCGCATTCGCAAGGCAAT..... 1060
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 eThrThrGluSerThrSerProAspSerSerThrGlySerThrSer 373
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1061 .....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
373 hrAlaGluProSerSerThrThrLeuThrProSerThrAlaThrPro 389
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1100 AATACCCGCTCCCTTACCATTCGCAAAATATCGTTCAACTTGAGCAG 1149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
390 SerThrAspGlnPheThrGlySerSerAlaSerThrGluSerAspSer 406
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1150 CGT.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
406 rAspSerSerThrValProThrThrGlyThrGluSerThrThrGluSer 423
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1181 CCGTGGCGCGCTCAACAGGAAAGATGTGAACCTGCGCAACCAAGCCAC 1230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
423 eSerThrThrThrAlaThrThrThrThrThrThrThrThrThrThrThr 437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1231 CCGAAGACCAAA..... 1242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
438 GluSerThrGluAlaLeuGluThrProAspGlyAsnThrThrSerGlyAs 454
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1243 ....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
454 nThrThrProSerProSerProAlaGlyThrProSerPheAlaAspThr 471
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1286 TAAATATGATACGAGAAATTAATACGCGCTTACCACAAAGATTCCTATA 1335
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
471 lThrProAspAsnGlyValSerThrGlnHisThrThrIleAsn..... 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1336 GATGAACCCGCTTTAATCCTAA..... 1359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
486 AspHisThrThrAlaAsnAlaGlnLysHisAlaGlyHisHisArgGlyArg 502
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1360 .....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
502 gAlaGlyArgGlyArgGlySerProGlnGlySerHisThrThrThr..... 517
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


A:Residues: 1-1609 <PRO>
A:Cross-references: EMBL:X59720; NID:g1907116; PID:e264634; PID:g1907227; MIPS:YCR089w
C:Genetics:
A:Gene: SGD:FIG2
A:Cross-references: SGD:S0000685; MIPS:YCR089w
A:Map position: 3R
C:Keywords: transmembrane protein
F:4-20/Domain: transmembrane #status predicted <TM1>
F:1592-1609/Domain: transmembrane #status predicted <TM2>

alignment_scores:	
Quality:	131.50
Ratio:	0.460
Percent Similarity:	50.530
	Length:
	Gaps:
	25
	Percent Identity:
	20.848

alignment_block:
MS-09-303-518D-465 v. 03E3AE

Align seg 1/1 to: S25345 from: 1 to: 1609

```

17  AATATACCCCTTATTTCTTCCACTGCGATGCGTCCGCGATGACGCA 66
   ::::: |||:::|||||
763  GluTyrThrThrTyrPro.....AlaSerSerIleAlaTy 775
   ::::: |||:::|||||
67  CACGCTCAGATT...TGGCAACGATCTTTATCCGCGAGTTTCGA 113
   ||| ::::: |||:::|||||:::|||||::: |||
775  rThrThrSerIleSerTyrLysThrIleuValIleThrThrGluValCys 792
   :: ||| ::::: |||:::||||| |||::: ::::: |||:::|||||
114  CGGTAGCGATTTCGAACCCGAGGGGAATACCCACTATTCGGCAGCAGG 163
   :: ||| ::::: |||:::||||| |||::: ::::: |||:::|||||
792  eHisSerLeuCysThrProThrValIleThrSerValThrAlaThrSer 808
   :: ||| ::::: |||:::||||| |||::: ::::: |||:::|||||
164  GGGAGCTGTCGC.....AGCCACGCGT 186
   ::::: |||
809  SerThrIleProLeuIleuSerThrSerSerSerThrValIleuSerSerTh 825
   ::::: |||
187  CATATCGATTGGGAACATACAAACCATCAGTTGGGCAACCTCTTCAT 236
   ::::: ||| ||| ||| ::::: |||
825  rValSerGluGlyAlaLysAsnProAlaIleSerGluValThrIleAsnT 842
   ::::: |||
237  CCAGCAGCGGGCCATTAAAGAAATATCGGCTACA.....TTGTCC 277
   ::::: ||| ::::: |||
842  hGlnValSerAlaThrSerGluAlaThrSerThrSerThrGlnValSer 858
   ::::: ||| ::::: |||
278  GCTTTTCGATACGCGGCGCGAAGTCATTCCTCCCTGCACAAACATGCG 327
   ||| |||::: |||::: |||
859  AlaThrSerAlaThrAlaThrAlaSer.....GluSerSerThrThrSer 873
   ||| |||::: |||::: |||
328  TCACATTCGCGATTCTGATGAAACCCGCTAGTCCCTGACAGCATTCAGCT 377
   ::::: ||| ::::: |||
873  rGlnValSerThrAlaSerGluThrIleSerThrIleuGlyThrGlnAsp 890
   ::::: ||| ::::: |||
378  TTACCGCATCCATTGGACGCGATACGAACACCATCCCGCGACGGCTATG 427
   ||| ||| ::::: |||
890  heThr.....ThrThrGlySerLeuIleuPheProAlaIleuSer 902
   ::::: |||
428  ACGGGCCACAGGGCGGGGCTATCCCGCTCCCAAGAGCGCGAGGATATA 477
   ||| |||::: |||
903  ThrGluMetIleAsnThrThrValValSerArgLysThrIleIleLeu 919
   ::::: |||
478  TACA.....GCTACGACATAAAGCGCTGGCCCAAAATATCCGGCTCA 521
   ||| ::::: |||::: |||::: ||| ::::: |||
919  rThrGluValLysSerThrSerLysCysValProThrValIleThrGluV 936
   :: ::::: |||::: |||::: |||
522  CCGTACGACAAACCGCACACCGCGAACAAGCGCTTGC..... 559
   :: ::::: |||::: |||::: |||
936  aValThrSerLysGlyThrProSerArgLysIleHisSerSerGlnThrIleu 952
   :: ::::: |||::: |||::: |||
560  .....ACGCTTCCACAAATACCGGTAATGTAATGCTG 588
   |||::: |||::: |||::: |||
953  GlnThrGluAlaValGluValThrIleuSerSerThrGlnThrValThrMe 969
   ::::: |||

```

[illegible]

